...OGICAL ROOM
...RAL LIBRARY
...Y. OF MICH.

NOV 9 1928

OCTOBER, 1928

Journal of Educational Research

CONTENTS

	Page
Measurement and Experimentation in Educational Methods — *V. A. C. Henmon*	185
The Retention of American History in the Junior High School — *Fowler D. Brooks and S. Janet Bassett*	195
The Vocabulary Load of Certain Secondary School Mathematics Textbooks — *H. H. Remmers and A. Grant*	203
A Study of a Modified Form of the Multiple Response Test — *Francis D. Curtis and Gerald G. Woods*	211
Unfamiliar Situations as a Difficulty in Solving Arithmetic Problems — *Carleton Washburne and Mabel Vogel Morphett*	220
A Table for Predicting the Validity and Reliability Coefficients of a Test When Lengthened — *Harold A. Edgerton and Herbert A. Toops*	225
A Comparison of Two Methods of Learning — *R. D. Russell*	235
An Experimental Investigation of the Relative Effectiveness of Two Plans of Supervised Study — *Harl R. Douglass*	239
Editorials	246
Reviews	249
Research Abstracts and Bibliographies	254
News and Communications	260

Copyright 1928, by Public School Publishing Co.

Entered as second-class matter at the Post Office at Bloomington, Illinois under the act of March 3, 1879

JOURNAL OF EDUCATIONAL RESEARCH

This journal was established to support educational research workers in all phases of the great problems of teaching. Started in 1920 under the editorship of Dr. B. R. Buckingham and associates among the members of an association of directors of educational research, and accepted as the official organ of the Educational Research Association, its purpose has at no time been confined to a narrow line of specialized research but has been concerned also with professional applications for all workers in education.

EDITORIAL BOARD

E. J. ASHBAUGH
Ohio State University

A. S. BARR, Chairman
University of Wisconsin

EARL HUDELSON
University of Minnesota

ASSOCIATE EDITORS

B. R. BUCKINGHAM, *Harvard University*

W. W. CHARTERS, *Director of Bureau of Educational Research, Ohio State University*

M. E. HAGGERTY, *Dean, College of Education, University of Minnesota*

V. A. C. HENMON, *Professor of Psychology, University of Wisconsin*

MURRAY DALMAN, *Educational Consultant, Perkins, Chatten and Hammond, Architects, Chicago, Illinois*

VIRGIL E. DICKSON, *University of California, Berkeley*

CHARLES FORDYCE, *Director, Bureau of Educational Measurements, University of Nebraska*

WILLIAM S. GRAY, *Dean, School of Education, University of Chicago*

KARL J. HOLZINGER, *School of Education, University of Chicago*

ERNEST HORN, *College of Education, University of Iowa*

ARTHUR W. KALLOM, *First Assistant Director, Department of Educational Investigation and Measurement, Public Schools, Boston*

WILLIAM A. McCALL, *Associate Professor of Education, Teachers College, Columbia University*

GEORGE D. STRAYER, *Teachers College, Columbia University*

W. L. UHL, *Dean, College of Education, University of Washington*

GUY M. WHIPPLE, *Secretary, National Society for the Study of Education*

F. J. KELLY, *President, University of Idaho*

EUGENE NIFENECKER, *Assistant Director, Bureau of Reference, Research and Statistics, Public Schools, New York City*

W. J. OSBURN, *College of Education, Ohio State University*

SHELTON PHELPS, *Director of Instruction, George Peabody College for Teachers*

J. B. SEARS, *Stanford University, California*

J. L. STENQUIST, *Director, Bureau of Educational Research, Public Schools, Baltimore, Maryland*

W. W. THEISEN, *Assistant Superintendent of Schools, Milwaukee, Wisconsin*

C. W. WASHBURNE, *Superintendent of Schools, Winnetka, Illinois*

CLIFFORD WOODY, *Director, Bureau of Tests and Measurements, University of Michigan*

THE JOURNAL OF EDUCATIONAL RESEARCH is published at Bloomington, Illinois, monthly, except July and August by the Public School Publishing Company. The subscription price is $4.00; $4.20 in Canada or Mexico; and $4.50 in other foreign countries.

Subscriptions and all communications about reprints or other business matters should be addressed to the PUBLIC SCHOOL PUBLISHING COMPANY, at Bloomington, Illinois.

Address all communications on editorial matters to A. S. Barr, Department of Education, University of Wisconsin, Madison, Wisconsin.

JOURNAL *of* EDUCATIONAL RESEARCH

Volume XVIII OCTOBER 1928 Number 3

MEASUREMENT AND EXPERIMENTATION IN EDUCATIONAL METHODS[1]

V. A. C. HENMON
University of Wisconsin

FOR three years I have been associated with the Modern Foreign Language Study, interested particularly in the production of prognostic and achievement tests as instruments of analysis and in the setting up of controlled experiments on objectives, curriculum-content, organization and method in the teaching of the modern languages. This work covering as it did necessarily a wide range of problems has occasioned many hours of reflection on the possibilities and also on the difficulties of measurement and experimentation in education. The outstanding characteristic of our generation in education is the application of quantitative and experimental methods. We have taken seriously the ideal of the exact sciences "to measure that which is measurable and to render measurable that which is not." The challenging axiom of one of our leaders, "Everything that exists at all exists in some quantity and can be measured," we have accepted, and the application of quantitative methods goes on apace. We teach our students to be scornful of tradition and of mere observation and insist that all things must be subjected to the test of controlled experimentation.

This is undoubtedly a healthy attitude to take if education is to become a science but the constant reader of present day educational literature cannot in his critical moments help but be troubled by the imperfections and ambiguities of our measurements and the inconclusiveness of our sporadic experiments. When, for example, on such an important problem for educational theory and practice as the effect of equal practice on individual differences, whether equal practice increases or decreases them, we find out of twenty-four experimental studies twelve of them leading at least tentatively to the conclusion that differences are increased and twelve to the conclusion that differences are decreased, we cannot help wondering about our experiments and about the conclusions

[1] Address before the *Eighth Annual Session of the Ohio Educational Conference.*

derived from them. Experimentalists in education perhaps need to be reminded of Aristotle's observation, "Those who take in but a few considerations easily decide." The temptation to play up apparently justified conclusions, especially when they run counter to generally accepted notions, is too strong to be resisted and is the bane of education in America. Bacon's criticism of experiment in his day applies aptly to much educational experiment in our day, "Even when men build any science and theory upon experiment, yet they almost always turn with premature and hasty zeal to practice, not merely on account of the advantage and benefit to be derived from it, but in order to seize upon some security in a new undertaking of their own not employing the remainder of their labor unprofitably, and by making themselves conspicuous, to acquire a greater name for their pursuit. Hence like Atalanta, they leave the course to pick up the golden apple, interrupting their speed, and giving up the victory." A more recent critic, not unsympathetic with measurement and experimentation, writes, "We live nowadays in American education amid a perfect orgy of statistical methodology, much of which can make but little or no contribution to constructive pedagogy. For those who expect something *Allcinseligmachend* from these proceedings one might paraphrase from Mephistopheles:

"Der Statistiker der tritt herein,
Und beweist euch, es musst so sein:
.....
Das preisen die Schüler aller Orten,
Sind aber keine Weber geworden."

For a dozen years or more we have sharply distinguished between intelligence tests as measures of capacity to do the things that lie ahead and achievement tests as measures of what lies behind and now Kelley in his recent book on the *Interpretation of Educational Measurements*, has shown that this is very largely if not entirely a spurious distinction and that the accomplishment quotient or ratio, of which so much was predicted, is an unsound measure. One cannot help but think of the lines from Browning's "Cleon" that Cattell is so fond of quoting:

"And I have written three books on the soul,
Proving absurd all written hitherto,
And putting us to ignorance again."

It may be pointed out in passing that the present almost chaotic condition in our conceptions of what psychology is about makes this quotation now peculiarly appropriate. Those who believe that psychology, if it is anything at all, is the science of the inner life are dubbed the medi-

cine men of a prescientific age and beyond the pale by those to whom psychology is a science of behavior. Those who believe that while "psychology long ago lost its soul, is now losing its mind, but can never quite lose consciousness," more politely dubs the extreme behaviorist as a biologist or as a physiologist. While the psychologist is likely to be scornful of education as a pseudo-science, he too has troubles of his own and finds his footing among biologists and physiologists pretty insecure.

Twenty-five or more years ago, when we began to apply systematically the method of correlation to the study of the relationships between mental traits, the early results showed a surprisingly low correlation. One of two interpretations was possible, either our measurements of the traits in question were imperfect or else there is a high degree of independence in mental functions. The latter interpretation was accepted and has for a quarter of a century influenced very materially our theory of curriculum-making and methods, and our notions on transference of training. The reader of the accumulating literature on this important problem for educational and vocational guidance and for educational theory and practice can, however, not escape the conclusion that with the improvement in statistical methods notably by the corrections for attenuation proposed by Spearman and methods of partial correlation, with the improvement in the reliability and validity of measurements, with the increase in the number of cases, control of or allowance for the variability of groups, and the equalization so far as possible of effects of practice, the correlations reported have steadily increased. Whatever may be the causes, whether due to a common central factor or to group factors, the relationships between traits are not only positive but high. Imperfections in our measurements not low relationship between mental traits appears from the accumulating evidence to have been the sounder interpretation. Our educational theorizing and practice have, however, been shot through and through and still are by notions of the extreme specialization of mental functions and the non-transferability of training. In a reaction against discipline we have made knowledges and skills sacrosanct. Karl Pearson maintained years ago that "the true aim of the teacher must be to impart an appreciation of method and not a knowledge of facts. This is far more readily achieved by concentrating the student's attention on a small range of phenomena, than by leading him in rapid and superficial survey over wide fields of knowledge." I believe in the soundness of both propositions but they run quite counter to the educational philosophy which has underlain much of our thinking on curriculum and

methods for two decades. The philosophy itself grew out of imperfect measurements and inconclusive experiments.

Since James made his experiments on the transfer of training, we have had upwards of forty experimental studies on the different aspects of this central problem of education. We are about as far from a solution as we were when we began. When the Committee on the Reorganization of Secondary Mathematics attempted to formulate some inferences which might reflect the present status of disciplinary values in education in the light of experiment, there resulted seven rather vague statements one of which was to the effect that "the true amount of transfer from one field to another has not been found by experiment, on account of one or more of the following handicaps:

(a) The maturity or previous training of the subjects tested.
(b) The absence from the training of the factors most favorable to transfer.
(c) The inadequacy of the tests to measure the traits sought.

We have a single important generalization out of the welter of experimentation and discussion, that there is transfer in so far as there are common or identical elements, and even this Judd characterizes as vague, ambiguous and useless. Judd has proposed a theory of generalization and in the literature we have two supposedly contrasting theories presented, the theory of identical elements and the theory of generalization, as if generalization could mean anything else than finding identities between experiences. It is easily conceivable that of two pupils seated side by side under the same teacher, one because of fertility of ideas, capacity to see similars in experiences, sagacity, mental alertness, or whatever we may call it, will secure transfer values, the other will not. It is also conceivable that of two teachers both good so far as any standards of technique of instruction we now possess are concerned, one will make instruction widely transferable, the other will not. Transfer then becomes a matter of individual differences in the learner and individual differences in teachers and teaching. Even if we grant that there is transfer only in so far as there are common or identical elements, who knows whether the elements common to two activities or two school subjects are many or few. One may claim that the common elements are many or important, the other that they are few or unimportant, without fear of successful contradition. We have not settled the quarrels about disciplinary values or transfer nor is there any prospect of our doing so. As a matter of fact Thorndike's important study of mental discipline in high-school subjects which goes to show that while the general

improvement due to studies in the capacity to think is small, the differences between studies in respect to it are likewise small, can very well be interpreted to mean great community of elements in school subjects.

For some months I have been coöperating with Professor Frank N. Freeman in a canvass of the literature of educational psychology now offered to teachers in an effort to determine what facts, laws or principles educational psychology has to offer as specific guides to practice. It is a rather disconcertingly difficult task. An example or two will suffice. Educational psychology might be expected to offer specific suggestions on the effect of different modes of presentation in memorizing such things as lists of words, whether they should be presented to the eye or to the ear, or to both senses simultaneously, and whether the words should be articulated at the time of presentation. This is a very specific and definite problem. Here is what a very competent text in psychology has to offer:

> "Much experimental work has been performed to determine the relative value of these various modes of presentation. No consistent results have been obtained. Some experimenters have found that articulation is an aid with the majority of subjects, while the opposite result has been obtained in other experiments. Some have found that visual presentation is better than auditory for most subjects, while auditory presentation is the more efficient in other experiments. Sometimes the material may be memorized more readily when it is presented to the eye and ear simultaneously, and sometimes better results are obtained when one sense organ is employed. Evidently the relative values of the several modes of presentation vary with the individual, and probably with other conditions as well.
>
> "The individual factor is probably due in large part to differences in training and habits of study. A person who is accustomed to articulate words while memorizing them will naturally be at a disadvantage when requested to inhibit this tendency, while a person who does not articulate will be distracted by the attempt to do so. A college student who organizes and memorizes the materials of a lecture while listening to it will naturally be able to memorize better with auditory than with visual presentation, while those students who are accustomed to the textbook method of instruction will do better with vision. It is thus obvious that any statement concerning the relative values of any two modes of presentation is worthless unless the subjects employed in the experiment are equally practiced in memorizing material in these two ways."

If this is an accurate summary of the statements that can fairly be made, and it is, after upwards of forty experimental studies, how can rules of practice or any helpful suggestions for method be extracted from them? Similar difficulties in securing clear cut statements for the guidance of practice are encountered on such matters as the distribution and length of practice periods, on learning by wholes and by parts, on rate

of reading in learning, on retroactive inhibition, and on fatigue, to mention only some notable examples. These illustrations, and they could be greatly extended, may suffice to point the moral, that our measurements are too imperfect to base a true science of education or a methodology upon them, that our experiments are too inadequately controlled, and that educational processes are so complex that our relatively simple experiments are incapable of revealing them. When we bear in mind that these experiments, cited for illustrative purposes, are for the most part laboratory experiments and not carried out in schools and under school conditions, the inconclusiveness of them is the more apparent and the difficulties of interpretation and application become all the greater.

Methods of instruction, we should all agree, must conform to the laws and principles under which the organism works most economically and effectively in acquiring its experiences. But it is just these facts, laws, and principles of human behavior that it is so difficult to formulate and which if they could be formulated would be the real foundations of method. The highly integrated forms of response even of the six-year old child are so complex as almost to defy analysis. Every educational process or product that has been studied quantitatively and experimentally turns out to be much more complex and intricate than we had supposed it to be. Witness the situation in the two subjects—reading and arithmetic—most assiduously cultivated by experimental education. Both bristle with unsolved problems. We are not clear as to what reading ability is nor can anyone trace out at all accurately the way in which our reading habits are built up. Darwin long ago observed that we do not know a thing until we know how it came to be and what it is. Now that psychology is giving so much critical study to the simplest reflexes and the methods of conditioning them in infants and the higher animals, perhaps we can hope for greater clearness as to the way in which higher orders of habits are integrated.

It is easy to be critical of our achievements and to set forth our deficiencies but it is better that we who are interested in making education a science and founding an art upon it should be clearly conscious of our limitations than that others should point them out for us. The sage observation of Socrates that he who knows nothing knows much in knowing that he knows nothing, has much of comfort in it. Courtis has recently rather pessimistically declared "We have no science worthy of the name, and no prospect of developing one so long as we refuse to face the inadequacy of our present results and the non-scientific character of our

many rationalizations. . . . . We have deceived ourselves into considering our progress to be on a higher level than it really is. The challenge of the opportunity before us is as vast as it is unperceived by the rank and file of those in education whose interest is to be scientific." If this is the situation, and I think it is, what can be done about it? We have no lack of faith in measurement and experiment. We need much better measurements and more carefully controlled experiments and should take steps to secure them. We have forty or more experimental or laboratory schools and at least one experimental college but Rugg and Counts pointed out two years ago that in not a single one of the laboratory schools was a controlled experiment being conducted. Our so-called experiments are rarely experiments in the scientific sense and seldom if ever yield results that are convincing to an unprejudiced observer. Such an observer at once sets about picking holes in the experiment and has no difficulty in doing so.

It may be urged and has been again and again since John Stuart Mill wrote on the logic of the moral and social sciences that experiment in the scientific sense is impossible in education. Thus Shorey maintains "In the physical sciences the experimental method isolates and discovers the true cause by elimination. The plurality of causes and the intermixture of effects preclude this procedure in the infinitely complex social sciences of ethology and education." To which we may reply with John Ray,—"Much may be done though we but endeavor and nothing is insuperable to pains and patience." The crude experimentation to which we now resort must, however, be replaced if we are to justify the experimental attitude and the experimental ideal in education.

The outstanding achievement in science during the last twenty-five years has been the development of genetics as an exact science. East in his recent remarkable book, *Heredity and Human Affairs* points out that, "Genetics was born and christened because of Gregor Mendel, not because he was such an intellectual giant that he could analyse and codify the complex results which had baffled his predecessors in hybridization work, but because he had the really brilliant idea of simplifying his experiments to the point where he was dealing with only one or two variables at a time. When heretofore botanists had crossed plants differing by hundreds of characters and had been bewildered at the apparent chaos of their data, Mendel used varieties which differed by a single striking character. This lone character he followed through generation after generation with the carefulness of a master workman, obtaining

results so simple that he was able to give them their correct interpretation." We need to learn this great lesson of controlling our variables better in educational experiment, and it cannot be too often emphasized, if we are to make real progress. We need to set about deliberately and consciously to create the means for this enterprise to control one by one the variables of importance. Problems of objectives to be realized and the conditions of their realizability, pupil capacities, curriculum content, methods of instruction, and standards of accomplishment are so complex and reveal so many uncontrollable factors or factors that can be controlled only with the greatest difficulty that only long-time, large-scale investigations will yield final and conclusive evidence. The difficulties of the case have been brought home to me very clearly in attempting to attack fruitfully two fundamental practical problems in the teaching of the modern languages—the age at which the study of a foreign language can most profitably be begun and how long the study must be continued to secure genuinely worth while results. These are problems on which there is such wide difference of opinion even among language teachers and such strong prejudices that only indubitable evidence has any chance at all of carrying conviction. The problems are just now of practical importance in connection with the developing junior high school and the place of the modern languages in its curriculum. The answer to the question of the age of beginning is so complicated by the purpose or aim of the student, his intellectual level, the length of time the study may be continued, the methods used and the social value of the modern languages in relation to the subjects they replace or by which they are replaced that convincing experiments are difficult to devise. At any rate only a carefully planned study extending over a ten-year period or longer could possibly tell the story for we are interested not only in immediate results but in the permanent outcomes after a lapse of time. Comparative studies of achievements of pupils beginning in the seventh, ninth, and eleventh grades and in the first year in college which show generally a higher achievement in such language capacities as can be measured as we ascend the age scale is a contribution to the problem but not a solution. Even if a language is learned with greater difficulty at the earlier levels, it may be a more permanent and usable acquisition when so learned.

The problem of method in the teaching of the languages has as every one knows been the subject of much controversy and acrimony since the days of Vietor. Discussion has centered about the relative merits of

the grammar-translation method and the direct method. To discuss methods on this level seems to me to lead nowhere. If a pupil expects to pursue the language for four years or more, one method may be the most economical and effective. If he studies the language for two years or less, as 84 percent of the pupils now do, another method may be most effective. The age of beginning, previous language training, objectives or purposes to be realized, the length of the course, and many other considerations determine method. Method will be among the last of the problems of education to be scientifically established. Until we are clearer about our objectives, content, and organization, we shall continue to flit from supervised study to socialized recitations, to problem-projects, to contract plans, and what not.

The only solution that I can see lies in the creation of research institutes in each of the school subjects for the direction of experiment and the dissemination of information in usable form. My fondest hope will be realized if out of the Modern Foreign Language Study there may grow a Modern Foreign Language Research Institute to carry on the needed investigations on which only a beginning has been made. We have reached the stage where we realize that we cannot solve our problems by theorizing or taking thought about them. We have also reached the stage when we find we can't solve anything by appointing committees and by taking votes. We have also, I hope, reached the stage when we cannot rely for progress on sporadic experiments here and there by interested teachers or on research by graduate students in our colleges and universities. Our great endowments could expend their resources to better advantage in supporting such institutes than in the support of short time committee investigations. If each of the schools of education in which research is being carried on would concentrate its special efforts on some one school subject and become the center of activity in this field in close coöperation with the schools of its state or region, we might control the numerous variables and we could expect results which we cannot under present conditions hope to attain. As matters stand now the enormous increases in secondary school and college enrollments have brought it to pass that those who are most capable of making fruitful investigations are swamped by teaching and administration and constructive research languishes.

I have said little about method in teaching and should perhaps apologize for the omission. The real foundation of method has still to be built. Method apart from objectives, content, and organization is an

abstraction without much meaning. All we can do is to dream of the day when principles of method that will stand the test can be formulated and in the meantime live as well as we can in our dreams. We are afflicted now with inadequate measurements, inconclusive experiments, and a realization of the complexity of our problems. We may take comfort that "our light afflictions, which are but for a moment, worked out for us a far more exceeding and eternal weight of glory."

THE RETENTION OF AMERICAN HISTORY IN THE JUNIOR HIGH SCHOOL[1]

FOWLER D. BROOKS AND S. JANET BASSETT
Johns Hopkins University

THE question of how well children retain the things they learn at school is of perennial interest. And, while many people believe that the schools are inefficient because children forget quickly so much of what they learn, yet there is no valid evidence showing just how permanent their learning really is. The majority of opinions are based upon uncritical observation or prejudice.

How well do pupils remember the history which they learn in junior high school? When do they forget most rapidly? Both of these questions are of practical importance for teaching. If we know the amount which pupils forget in a given time, we have a factual basis for evaluating the efficiency of instruction and can know if improvement is needed. If we know when pupils forget most rapidly, we have a basis for determining when reviews should be given. Current psychological opinion is based largely upon the classic work of Ebbinghaus on learning and relearning nonsense syllables, and upon other allied investigations. A much better basis is secured by ascertaining the amount of a subject known by pupils at a specified time and at stated intervals thereafter. Only in this way can adequate information be secured on the amount forgotten and on the rate of forgetting.

Two early attempts to find a solid basis of fact for this problem may be cited. Along in the spring of 1900 a committee was appointed to investigate (survey!) the public schools of the District of Columbia. The chief examiner of the Civil Service Commission prepared questions on the arithmetic, history, etc. taught in the grammar grades. These questions were given to 1188 first-year high-school pupils nine months after they had finished the grammar school. The arithmetic scores averaged 58.8 percent and the history scores 53.1 percent. The investigation committee concluded that nine months after leaving the grades more than 40 percent of the grammar school content could not be *recalled*,[2] although

[1] The material of this report is part of the Ph. D. dissertation of S. Janet Bassett at Johns Hopkins University, 1928. Other parts relate to sex differences in retention, factors influencing retention, and the kinds of history knowledge most readily forgotten and the kinds retained best.

[2] *Supplemental Report, School Examinations*, Apr. 14, 1900, Senate Report 711, part 2, 56th. Congress, First Session, 15 pp.

there was no evidence to warrant the assumption that the pupils ever did know all the things called for by the questions or ever could have made perfect scores on them; in fact, we are certain that these pupils never were able at any time during their grammar-school days to make average scores of 100 percent on the tests given.

A little later another investigation[3] of the amount of subject matter recollected by 1362 high-school students from one to five years after it had been studied shows that only 5 of 33 questions on history and geography were correctly answered by as many as 75 percent of the pupils. Here again, no information was available to show the amount which the pupils did know at any specified time previous to giving the tests to determine how much was remembered.

Recent investigations[4] of the permanence of school learning (excluding, of course, those to ascertain the effect of the summer vacation) do not have data showing the amount which students actually did know when they quit studying a subject at school, but infer the amount from other sorts of information, such as the scores made by other similar groups of pupils who have just finished the subject.

The present report shows the amount of American History actually retained by 495 students in Grades VII and VIII of several public schools in Baltimore from four to sixteen months after the semester in which it was studied.[5] The results indicate that pupils forget 23 percent of first- or second-year junior high-school American History in a year's time.

The Baltimore city course of study in American History in each half year of Grades VII and VIII is a distinct unit with no overlapping. Standardized history tests could not be used to measure retention, because they do not cover fully the subject matter actually taught in each half year. Accordingly, four comprehensive new-type examinations were prepared to cover the half-years, VIIB, VIIA, VIIIB, and VIIIA, respectively. Preliminary try-outs were made upon 187 pupils not otherwise

---

[3] Cober, E. W. "A Study of High-School Pupils with a View of Determining the Extent of Recollection of Once Familiar Facts." Thesis, University of Pennsylvania, 1912. 46 pp.

[4] For example, Eikenberry, D. H. "Permanence of High School Learning." *Journal of Educational Psychology*, Vol. XIV (1923), pp. 463-481; Powers, S. R. *A Diagnostic Study of the Subject Matter of High School Chemistry.* (New York: Teachers College, Columbia University, 1924), pp. 49-54; Thorndike, E. L. "The Permanence of School Learning." *School and Society*, Vol. XV (1922), pp. 625-627.

[5] Investigations on the retention of other subjects have been under way for two or three years by graduate students at Johns Hopkins University.

used in the investigation. The tests were revised in the light of the results from these try-outs and upon the combined judgments of history teachers and supervisors and several competent students of experimental education. In their final form, each of the tests consists of four parts,—40 true-false statements, 10 completion items, 20 multiple-choice elements, and 4 modified essay questions. On the latter, the student selected any three of the four questions and gave four statements to answer each of the three selected. A perfect score on each test is 82.

A part of the test used in Grade VIIA is given to indicate the nature of the tests.

AMERICAN HISTORY TEST, GRADE VIIA.

Copyright, 1928 by S. JANET BASSETT

PART I

Directions: Some of the statements below are true and some of them are false. Put + on the line at the left of a statement if the statement is true. Put — on the line at the left of a statement if it is not true. Go through the list of statements quickly, answering those of which you are sure, then go back and study the more difficult statements. Answer these if you think you know them.

1. Sample.

..... 3. Hamilton strongly objected to Jay's Treaty.

..... 4. For a long time the Barbary pirates in northern Africa preyed on the commerce of Christian nations.

..... 6. The Napoleonic Wars stopped all American commerce.

..... 7. One of the causes of the War of 1812 was that the British searched American vessels.

.....15. Spain sold us Louisiana.

.....16. We wanted Louisiana in order to secure an outlet to the Pacific Ocean.

.....17. Florida was explored and settled by Spain.

.....18. Florida was used as a naval base by England in 1812.

.....20. Mexico granted large tracts of land in Texas to settlers from the United States.

.....28. The West was rapidly settled after the discovery of gold in California.

.....29. Few immigrants have gone beyond the Appalachians.

.....30. The Indians of the Oregon region welcomed the White Man.

.....34. The acquisition of so much land in the West helped to do away with the slavery dispute.

.....39. Before the Civil War there was a lack of public schools in the Southern States.

.....41. The Erie Canal stopped the trade between the West and the South.

PART II

Directions: Some important words have been omitted from the following statements. Write in the blank the word that best completes the meaning of the sentence. Each blank represents one word.

1. Sample.
3. The hero of New Orleans was.....
4. The two kinds of tariff are.....and.....
8. The "Pathfinder" was .....
9. Mexico claimed that the boundary line between Mexico and Texas was the ....
11. The industrial growth of the ..... was largely dependent upon "King Cotton."

PART III.

Directions: In each sentence select the one word or phrase which you think makes the statement true. Underline this answer.

1. Sample.
2. England seized our sailors for her ships on the ground that they were (a) deserters from the English army; (b) good fighters; (c) subjects of England; (d) enemies of England.
4. "Don't give up the ship," was said by (a) Hull; (b) Jones; (c) Decatur; (d) Lawrence.
7. Tariff is (a) a veto; (b) a debate; (c) an amendment; (d) money paid on imported goods.
10. An important effect of the Louisiana Purchase was that it (a) helped Jefferson to continue his doctrine of "strict" construction; (b) made Jefferson very unpopular; (c) gained the friendship of Spain; (d) gave the United States control of the Mississippi.
16. One of the great hindrances to western development was the (a) invention of the steamboat; (b) danger from the Indians; (c) increase in immigration; (d) discovery of gold.
17. The last frontier was (a) along the Pacific coast; (b) along the Gulf of Mexico; (c) on the banks of the Mississippi; (d) at the eastern base of the Rockies.
21. A great American writer was (a) Patrick Henry; (b) Henry Hudson; (c) Dred Scott; (d) Henry Wadsworth Longfellow.

PART IV.

Directions: Select any three of the following problems about which you know the most important facts. Write four statements on each subject in the space which is left beneath the topic. Think before you write.

1. How did we gain our freedom on the sea?
2. What was the tariff policy of the United States in 1860?
3. Why was the Louisiana Purchase the greatest event in Jefferson's administration?
4. How has the Industrial Revolution improved American life?

By repeating the tests with 72 pupils, not otherwise used in the investigation, the practice effect was found to be small, and the reliabilities were found to range from .64 to .90 when the tests were repeated at intervals of one week to eight weeks. The pupils had taken a great many standardized tests and new-type examinations, so that we would expect little practice effect from repeating these tests, especially at intervals of four or more months.

The tests, in final form were given according to the following schedule. The VIIB test to the VIIB grades at the close of the semester and thereafter at intervals of four, eight, twelve, and sixteen months. The VIIA test to the VIIA grades at the close of the semester (and thereafter at intervals of four, eight, twelve, and sixteen months.) And similarly in Grades VIIIB and VIIIA. Thus we have data showing how much of the history taught during a semester was known at its close, and how

much was known at intervals of four, eight, twelve, and sixteen months thereafter. The results are shown in Tables I and II.

TABLE I

AMERICAN HISTORY SCORES IN GRADES VIIB, VIIA, VIIIB, AND VIIIA AT THE CLOSE OF THE TERM IN WHICH IT WAS STUDIED AND AT INTERVALS OF FOUR, EIGHT, TWELVE, AND SIXTEEN MONTHS THEREAFTER

GRADE	NUMBER OF PUPILS		At close of term	AMERICAN HISTORY SCORES After intervals of			
				4 mo.	8 mo.	12 mo.	16 mo.
VIIB	167	Mean	50.2	43.7			
		S. D.	13.2	11.5			
	107[a]	Mean	55.1	47.3	43.8		
		S. D.	12.5	10.9	10.6		
	64[a]	Mean	56.6	47.6	44.7	40.7	39.6
		S. D.	13.9	11.8	11.5	11.3	11.2
VIIA	142	Mean	59.9	53.8	50.2		
		S. D.	9.6	10.2	9.7		
	120[b]	Mean	61.1	54.4	50.6	48.6	
		S. D.	9.8	10.7	10.2	10.6	
	86	Mean	60.1		51.5		
		S. D.	9.5		8.9		
VIIIB	79	Mean	61.9	50.5			
		S. D.	10.4	11.2			
	44[c]	Mean	58.4	47.3	44.1		
		S. D.	11.2	12.2	12.6		
VIIIA	21	Mean	54.7	49.5	46.3		
		S. D.	6.3	6.7	6.3		

[a]Included in the group of 167 pupils of Grade VIIB who were retested only once; that is, after an interval of 4 months.
[b]Included in group of 142 pupils of Grade VIIA who were retested after intervals of 4 and 8 months.
[c]Included in group of 79 VIIIB pupils who were retested after an interval of 4 months.

TABLE II

PERCENTAGE OF AMERICAN HISTORY KNOWN AT THE END OF THE SEMESTER (IN WHICH IT WAS STUDIED) WHICH IS FORGOTTEN FOUR, EIGHT, TWELVE, AND SIXTEEN MONTHS THEREAFTER (GRADES VIIB TO VIIIA)

GRADE	NUMBER OF PUPILS	PERCENTAGE OF HISTORY FORGOTTEN IN			
		4 mo.	8 mo.	12 mo.	16 mo.
VIIB	167	12.9			
VIIB	107[a]	14.3	20.5		
VIIB	64[a]	15.9	20.9	28.0	31.0
VIIA	142	10.2	16.3		
VIIA	120[b]	11.1	17.2	20.5	
VIIA	86		14.3		
VIIIB	79	18.4			
VIIIB	44[c]	19.0	24.5		
VIIIA	21	9.6	15.4		

[a, b, c] See corresponding footnotes to Table I.

An examination of Tables I to III shows that forgetting is most rapid during the first four months after the close of the term in which the history is learned, and that thereafter it is much slower. Data on other individuals indicate that pupils at the beginning of a semester knew very little of the history which was taught during that term, their scores being, on the average, approximately 10 points on the appropriate tests

TABLE III

PERCENTAGE OF AMERICAN HISTORY KNOWN AT THE END OF THE SEMESTER (IN WHICH IT WAS STUDIED) WHICH IS FORGOTTEN FOUR, EIGHT, TWELVE, AND SIXTEEN MONTHS THEREAFTER (DATA COMBINED FROM GRADES VIIB, VIIA, VIIIB, AND VIIIA)

NUMBER OF PUPILS	PERCENTAGE FORGOTTEN AFTER			
	4 mo.	8 mo.	12 mo.	16 mo.
409	12.9			
314	12.8	18.8		
184	12.8	18.5	23.1	
64	15.9	20.9	28.0	31.0

Table III is to be read as follows: 409 pupils in Grades VII and VIII forget 12.9 percent of the history which they knew at the close of the term in which they had studied it; 314 pupils forget 12.8 percent in 4 months and 18.8 percent in 8 months; etc.

devised and used in this investigation. The results are approximately as follows: one-eighth is forgotten in four months, one-sixth in eight months, one-fourth in twelve months, and one-third in sixteen months.

The pupils who made high scores at the end of the term usually made high scores four, eight, twelve, or sixteen months later, and those who made the lower scores the first time they were tested usually made the lower scores at the subsequent retestings, as indicated by the correlations between scores at the end of the semester and those at succeeding intervals (Table IV), the coefficients ranging from .74 to .89 and averaging more than .80. Furthermore, those who rank high at the end of four months are likely to rank high at the end of eight months, the correlations ranging from .90 to .97 and averaging well over .90. This is not

TABLE IV

CORRELATIONS (PEARSON) BETWEEN AMOUNTS OF AMERICAN HISTORY KNOWN AT THE END OF A TERM AND THE AMOUNT KNOWN FOUR, EIGHT, TWELVE, AND SIXTEEN MONTHS THEREAFTER (GRADES VII AND VIII)

GRADE	NUMBER OF PUPILS	INTERVAL				
		0–4[a]	0–8	0–12	0–16	4–8
VIIB	167	.887				
VIIB	107[b]	.875	.843			.914
VIIB	64[a]	.891	.869	.811	.807	.953
VIIA	142	.859	.787			.909
VIIA	120[b]	.878	.826	.812		.934
VIIA	86		.834			
VIIIB	79	.804				
VIIIB	44[a]	.775	.768			.974
VIIIA	21	.775	.740			.899

[a] 0 means at close of term in which the history tested was studied.
[a, b, c] See corresponding footnotes to Table I.

Table IV is to be read as follows: The correlations between the history scores of 167 pupils in Grade VIIB at the close of the term and their scores 4 months later is .887, etc.

surprising; in fact, it is exactly what we would expect, since the most rapid forgetting takes place during the first four months after ceasing the study of any semester's work in the subject.

While the results of this investigation are in agreement with the widely accepted view that those who know the most also retain the most,

as shown by average correlations of more than .80, yet they are not in agreement on another assumed correlative of the foregoing proposition, because further analysis indicates that the converse is not necessarily

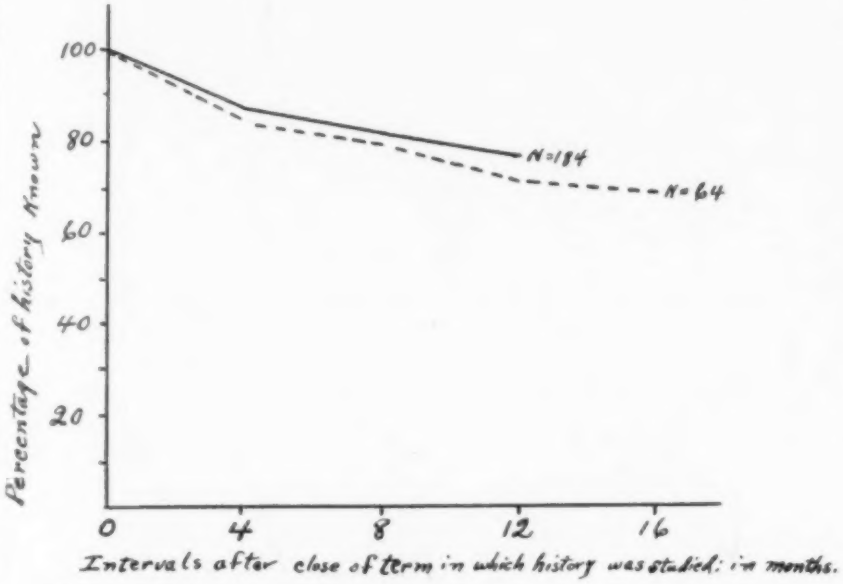

Fig. 1. Curves of Forgetting for American History in the Junior High School

true,—that is, that those who know the least forget the most. The first statement does not necessarily imply the second. As a matter of fact, the data of this study indicate that the pupils who know most at the end of the term tend to forget a little more during the ensuing months than those who make the lower scores at the first testing. The correlations between the scores at the first testing and the amounts actually forgotten during the succeeding months range from .022 (for 120 pupils in Grade VIIA retested twelve months after the initial testing) to .604 (for 64 VIIB pupils retested after sixteen months), they average .36 (Table V). There is no contradiction here. The pupils who know most at the end of the semester tend to forget a little more than those who know less, but they still know the most, preserving very much the same relative standings as on the initial test.

TABLE V

CORRELATION BETWEEN HISTORY SCORES AT THE CLOSE OF THE TERM AND THE AMOUNTS OF HISTORY FORGOTTEN DURING THE ENSUING FOUR, EIGHT, TWELVE, AND SIXTEEN MONTHS (GRADES VII AND VIII)

GRADE	NUMBER OF PUPILS	PEARSON CORRELATIONS BETWEEN SCORES AT CLOSE OF TERM AND AMOUNTS FORGOTTEN DURING INTERVALS OF			
		4 mo.	8 mo.	12 mo.	16 mo.
VIIB	167	.520			
VIIB	107[a]	.523	.526		
VIIB	64[a]	.529	.486	.589	.604
VIIA	142	.156	.291		
VIIA	120[b]	.086	.241	.022	
VIIA	86		.389		
VIIIB	79	.200			
VIIIB	44[c]	.223	.185		
VIIIA	21	.249	.352		

[a, b, c] See corresponding footnotes to Table I.

CONCLUSIONS

From the data of this investigation the following conclusions may be drawn:

1. In a year's time junior high-school pupils forget 23 percent of the American History which they knew at the close of the semester during which it was studied.

2. Forgetting is twice as rapid during the first four months after the semester in which the history material was presented as it is during the next four months; and it is slower thereafter.

3. Accordingly, reviews should be continued at least during the early part of the semester following the one in which the material is presented.

4. Pupils who know most tend also to retain most at later intervals, individual ranks being quite similar at the initial testing and the later retestings.

5. Pupils who know most tend also to forget a little more than those who know less, but they still maintain relative standings at later retestings quite similar to those on the initial ones, the average correlations being more than .80.

6. Accordingly, the best students need reviews as much as do the average or poorer students.

7. After the fourth month of forgetting pupils make little change in relative standings, all the correlations being .90 or more.

8. Whenever courses of study in the junior high school provide separate and distinct units of work for each half year, definite provision is needed for reviewing each term's work during the following term. In this way permanence of learning can be more effectively secured.

THE VOCABULARY LOAD OF CERTAIN SECONDARY SCHOOL MATHEMATICS TEXTBOOKS

H. H. REMMERS and A. GRANT
Purdue University

PURPOSE

THE purpose of this study was twofold: (1) to measure the vocabulary burden of secondary mathematics textbooks, and (2) to determine the reliability of measurement in the methods employed. It is assumed that the ease with which the pupil acquires the concepts of a subject is a function of the textbook's vocabulary load, and that therefore a measurement of this load would constitute an expository device of value in selecting a textbook.

METHOD

The method used was for the most part identical with that used by Bertha A. Lively and S. L. Pressey in their work on "A Method for Measuring the Vocabulary Burden of Textbooks."[1] Such variations as have been introduced will be noted and the reason for their use explained.

The first problem was to secure such sampling of words as would warrant statistical inference relating to the entire book. One thousand words were taken as probably adequate to constitute a sampling, and also as convenient to work with. To obtain such a sampling, the number of pages in the book was first noted, with the approximate number of words per line; the number of pages which should be sampled, taking one line on each page, was then found. The pages were so chosen as to insure even distribution throughout the book. All the words in these lines were then gone over and the different words listed alphabetically.

The thousand words having thus been noted, the total number of different words per thousand was then found. This gave the vocabulary range. These words were then looked up in the Thorndike *Word Book*,[2]

[1] *Educational Administration and Supervision.* Volume IX, October, 1923.

[2] Thorndike, E. L. *The Teacher's Word Book.* New York City: Teachers' College, Bureau of Publications, Columbia University. This book lists the ten thousand most common words of the English language, as determined on the basis of an elaborate investigation by Professor Thorndike. In the *Word Book* each word is followed by an index number indicative of its commonness. Thus such a common word as *and* has an index number of 210; a relatively uncommon word like *atom* has an index number of 4; still more rare words such as *neolithic* do not appear in the word book at all—those are listed in this study as zero value words. Words with credit numbers of 49 or over occur in the first thousand words, in frequency. Words with index numbers of 10 or over occur in the first five thousand words—and so on; for a more detailed statement, the reader is referred to the *Word Book* itself.

and the index number of each found. This number indicates the commonness of the word as found by Thorndike. The number of words not appearing in Thorndike's *Word Book* was then listed. Since these do not appear, they are not among the ten thousand most common words, and can be regarded as having zero for their index value, and as constituting the highly technical part of the vocabulary. The median index number was then computed. This is merely a numerical index of the book's vocabulary difficulty. Evidently, *the lower the index number, the more difficult the book.* (See Fig. 3). Here it should be noted that in computing the index number, Lively and Pressey counted the zero valued words twice. However, since this does not change the relation between index numbers, we considered this unnecessary.

Twelve different books were thus studied, as follows: Avery, Royal A. *Plane Geometry.* (New York: Allyn and Bacon, 1925); Hart, Clara A., and Feldman, Daniel D. *Plane Geometry.* (New York: American Book Co., 1911); Smith, D. E. *Essentials of Plane Geometry.* (Boston and Chicago: Ginn and Co., 1923); Schultze, Arthur and Sevenoak, F. L. *Plane and Solid Geometry.* (New York: Macmillan Co., 1907); Newell, M. J., and Harper, G. A. *Plane and Solid Geometry.* (Row, Peterson & Co., 1915); Wentworth, George and Smith D. E. *Plane and Solid Geometry.* (Boston and Chicago: Ginn and Co., 1913); Ford, W. B., and Ammerman, Charles. *First Course in Algebra.* (New York: Macmillan Co., 1921); Ford, W. B., and Ammerman, Charles. *Second Course in Algebra.* (New York: Macmillan Co., 1920); Wells, Webster. *Algebra for Secondary Schools.* (Boston and Chicago: Heath & Co., 1908); Young, J. W. B., and Jackson, L. L. *Elementary Algebra First and Second Course.* (New York: Appleton & Co., 1910); Wells, Webster and Hart, Walter W. *New High School Algebra.* (Boston and Chicago: Heath & Co., 1912); and Smith, D. E., and Reeve, W. D. *Essentials of Algebra.* (Boston and Chicago: Ginn and Co., 1924).

Tables I, II and III summarize the findings. For each book two counts were made to compare the observed with the theoretical variability. The method used for the second count was identical with that for the first count, except that a different page was used. The probable errors were found for the vocabulary range and for the zero words. In both cases this is done by use of the formula[3] $P.E. = .6745\sqrt{\frac{PQ}{N}}$ where

[3] Kelley, T. L. *Statistical Method.* Pages 89-90, on the standard deviation of a proportion.

N is the number of words (1,000), P is the number of successes (i. e., in this case different words), and Q the number of failures (i. e., N minus P). The application of the formula may be illustrated with the first count from Smith (Table I): P = probability of success (220);

Q = probability of failures (1,000 minus 220); N = 1,000;

$$\text{P.E.} = .6745\sqrt{\frac{220 \times 780}{1{,}000}} = 8.82.$$

TABLE I

NUMBER OF DIFFERENT WORDS FOUND ON TWO COUNTS IN CERTAIN SECONDARY SCHOOL TEXTBOOKS IN MATHEMATICS

	First Count	P.E.	Second Count	P.E.	Diff.	P.E. diff.	Diff. / P.E. diff.
Smith	220	8.82	225	8.88	5	12.48	0.40
Avery	234	9.02	243	9.14	9	12.83	0.72
Hart & Feldman	248	9.21	236	9.02	12	12.88	0.93
Newell & Harper	260	9.34	272	9.51	12	13.30	0.90
Wentworth & Smith	274	9.52	283	9.55	9	13.46	0.67
Schultze & Sevenoak	301	9.76	312	10.02	11	13.95	0.79
Smith & Reeve	251	9.00	270	9.49	19	13.07	1.45
Ford & Ammerman I	277	9.52	290	9.66	13	13.54	0.96
Ford & Ammerman II	320	9.96	332	10.06	12	14.11	0.85
Wells & Hart	300	9.76	325	10.00	25	13.95	1.79
Wells	342	10.10	354	10.18	12	14.33	0.80
Young & Jackson	380	10.38	362	10.24	18	14.58	1.23

This table is to be read as follows: The first 1,000 word count for Smith yielded 220 ± 8.82 different words; the second count, 225 ± 8.88 different words. The difference of the counts is 5 ± 12.48, which is statistically not significant.

TABLE II

ZERO WORDS

	First Count	S.D.	Second Count	S.D.	Diff.	P.E. diff.	Diff. / P.E. diff.
Smith	23	4.54	25	4.70	2	4.40	0.45
Avery	33	5.32	34	5.41	1	5.12	0.20
Hart & Feldman	34	5.41	27	4.90	7	4.92	1.42
Newell & Harper	56	6.62	45	6.13	11	6.08	1.62
Wentworth & Smith	54	6.57	57	6.72	3	6.33	0.47
Schultze & Sevenoak	51	6.50	62	7.05	11	6.45	1.71
Smith & Reeve	40	5.79	48	6.27	8	5.75	1.39
Ford & Ammerman I	36	5.58	36	5.61	0	5.33	0.00
Ford & Ammerman II	31	5.21	45	7.25	14	6.01	2.34
Wells & Hart	40	5.86	48	6.39	8	5.85	1.37
Wells	49	6.59	54	6.76	5	6.36	0.79
Young & Jackson	46	5.31	49	6.50	3	5.65	0.53

This table is read in the same way as Table I.

This result is to be interpreted to mean that with further samplings from this book it is an even wager that on the average the values for the category of "different words" will go as high as 228.82 and as low as 211.18. This application of the theory of sampling constitutes an important check on the reliability of the counts, and hence on the method used.

RESULTS

Range of Vocabulary.—The results are best shown in graphic form. Figure 1 shows the difference in range. It is evident that books vary widely, 220 to 380 different words per thousand, with respect to each other. The range within books is, however, small when compared with ranges found in other familiar books, by Lively and Pressey. For ex-

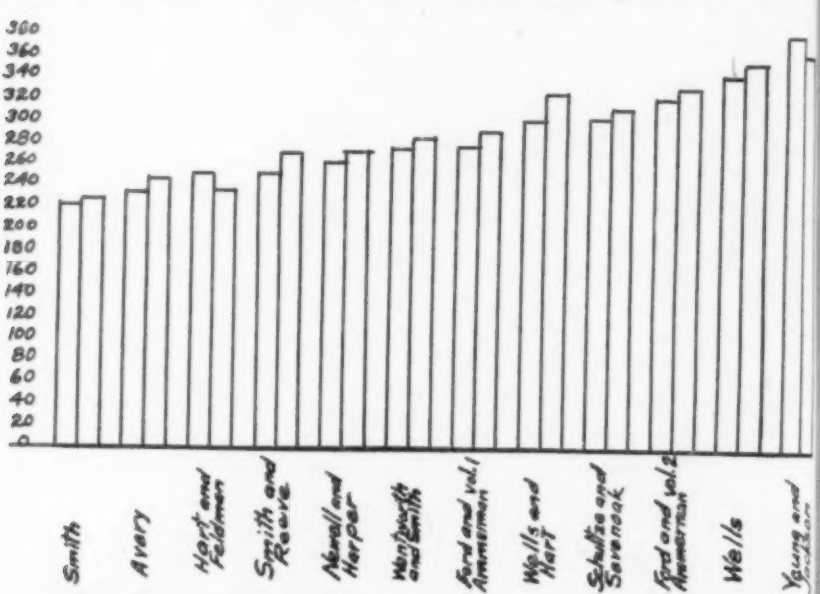

Fig 1. The number of different words for two samplings of 1000 words for each book.

ample, Muzzy's *American History* has a range of 533; Thackeray's *Vanity Fair,* 490; and Horace Mann's *Second Reader,* 412. It seems that the most difficult vocabulary burden in the books investigated with respect to range does not equal that of Mann's *Second Reader.* Evidently the same words are used over and over again, and the vocabulary used is not large.

Number of Zero Words.—Here again graphical representation is used. The range in this case is also wide, being 23—62. Comparison with the results found by Lively and Pressey are again interesting. Muzzy's *American History* has 27 zero words, and McPherson and Henderson's *General Chemistry* has 68. This latter number is higher than the number of zero words in the most difficult of the mathematics books investi-

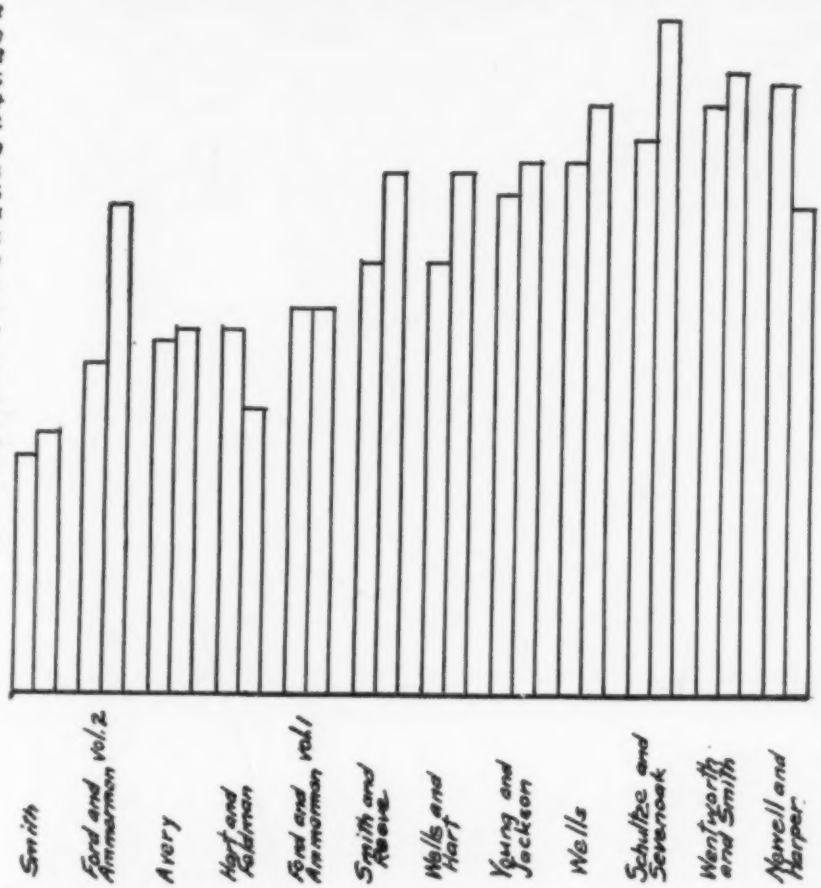

Fig. 2. The number of zero words for two samplings of 1000 words for each book.

gated. It is evident that the number of technical words in some other high-school books is higher than that found in the mathematics books. Evidently the same words are used often.

Since the number of zero words represents the size of the technical vocabulary, these numbers may be used as a basis of comparison. Before this is done, however, it is necessary to take into account the errors which each involves. Since these measures are unrelated, the probable error of the difference of two averages, due to variable errors of measurement and sampling, is given by the formula:[4]

[4] Kelley, T. L. *Statistical Method.* Page 182 on the standard deviation of a difference.

$$P.E._{Difference} = \sqrt{P.E.^2_{average_I} + P.E.^2_{average_{II}}}.$$

With these computed it is possible to compare any two books. Comparisons were made among different geometry books, different algebra books, and between geometry and algebra books in general. These comparisons are listed in Tables II and III. It will be noted that since the number of zero words is small, the P. E. is correspondingly large.

TABLE III

INTER-COMPARISONS OF VARIOUS TEXTBOOKS IN GEOMETRY

Textbook	Zero Words	S.D.	Diff.	P.E.$_{diff.}$	$\frac{Diff.}{P.E._{diff.}}$
Smith	23	4.54	10	4.71	2.12
Avery	33	5.32			
Smith	23	4.54	11	4.76	2.31
Hart & Feldman	34	5.41			
Newell & Harper	56	6.62	33	5.31	6.22
Smith	23	4.54			
Smith	23	4.54	28	5.34	5.25
Schultse & Sevenoak	51	6.50			
Smith	23	4.54	31	5.38	5.76
Wentworth & Smith	54	6.57			
Avery	33	5.32	21	5.70	3.68
Wentworth & Smith	54	6.57			
Avery	33	5.32	23	5.72	4.03
Newell & Harper	56	6.62			
Avery	33	5.32	18	5.66	3.18
Schultse & Sevenoak	51	6.50			
Hart & Feldman	34	5.41	22	5.76	3.82
Newell & Harper	56	6.62			
Hart & Feldman	34	5.41	20	5.74	3.48
Wentworth & Smith	54	6.57			
Hart & Feldman	34	5.41	17	5.70	2.98
Schultse & Sevenoak	51	6.50			

This table is to be read: The first 1,000 word count of Smith yielded 23 zero words with a S.D. of 4.54. Avery yielded 33 with a S.D. of 5.32. The difference is 10 ± 4.71. This indicates no significant difference in the difficulty of the two books.

Assuming that three probable errors constitute a significant difference, the following conclusions may be drawn:

AMONG GEOMETRY BOOKS.

1. Smith, Avery, and Hart and Feldman all have a significantly smaller technical vocabulary than Schultze and Sevenoak, Wentworth and Smith, and Newell and Harper.
2. The largest difference exists between Smith and Wentworth and Smith, and between Smith and Newell and Harper.

AMONG ALGEBRA BOOKS.

1. There is no real, significant difference between any of the algebra books.

AMONG GEOMETEY AND ALGEBRA BOOKS.

1. All geometry books investigated have a larger technical vocabulary load than Smith's *Algebra*.
2. Avery has a smaller load than Young and Jackson, or Wells.

Median Index Number.—These numbers are shown in Figure 3. Schultze and Sevenoak's Geometry again shows up as the most difficult, having an index number of 24. Avery's Geometry has the lightest load,

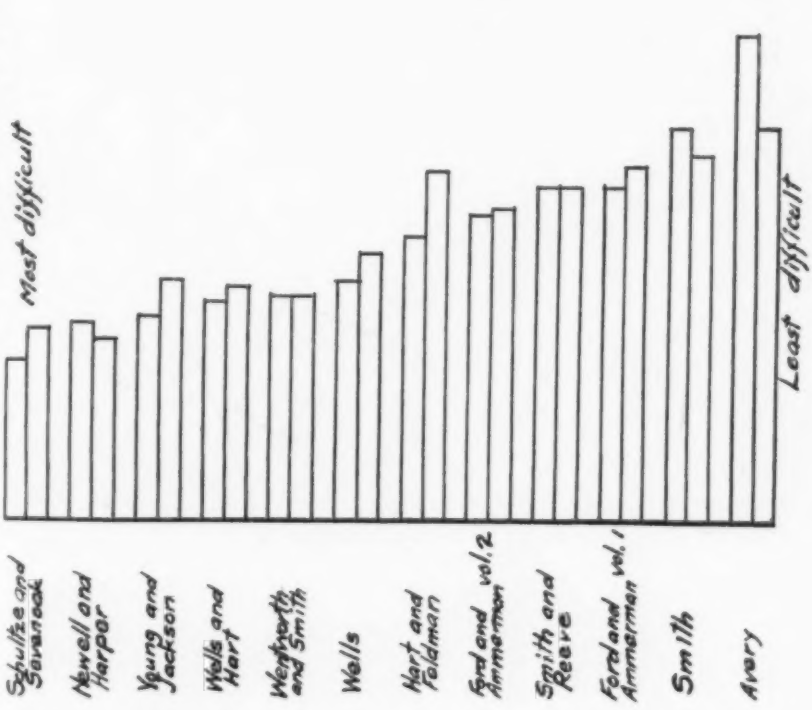

Fig. 3. Median difficulty indices (Thorndike) for two samplings of 1000 words for each book.

having an index number of about 60. It will be noted again that the lower the index number, the more difficult the book's vocabulary load.

Reliability of the Method Used.—The P. E.'s of the various categories studied, the number of different words, the zero word differences among different books and differences between two consecutive counts indicate

the variability expected on the basis of the theory of sampling. A comparison of these values with the observed values in terms of the ratio $\frac{\text{Diff}}{\text{P. E.}_{\text{diff}}}$ indicates that the theoretical and empirical data agree. It will be noted from Table IV that the rank-order correlation between median index numbers and number of zero words (both averaged for the two counts) is .81.

TABLE IV

RANK-ORDER CORRELATIONS FOR THE CATEGORIES INDICATED

	ZERO WORDS	MEDIAN DIFFICULTY
Zero word		
Median difficulty	.81	
Different words	.54	.60

This warrants the statement that either criterion may be used without danger of serious error. None of the shifts in rank are particularly significant as may be seen from an inspectional comparison of Figures 2 and 3. The correlations for the number of different words with the number of zero words and the median difficulty, while both positive and significant, are not high enough to warrant the conclusion that they may be used interchangeably as measures of vocabulary load.

A STUDY OF A MODIFIED FORM OF THE MULTIPLE-RESPONSE TEST

FRANCIS D. CURTIS
University of Michigan
and
GERALD G. WOODS
University of Michigan High School

THE purpose of this investigation was to determine some of the relative merits of the conventional multiple-response test having five choices, as compared with a modified form of this test. This modified form consists of the conventional form with a blank added after each exercise. The pupils are instructed to underline the correct response if it is present among those given, or, if the correct response is not given, to write it in the blank provided.

METHOD

Two mimeographed sheets of test items were prepared for each class. Sheet I contained seventy-five items of the conventional multiple-response type having five choices. Sheet II contained sixty of the same seventy-five items which appeared on Sheet I, and fifteen additional items. The sixty were unchanged in wording, but to make them seem different from the same items on Sheet I, their order and that of the five responses in each item was changed; each, moreover, was made into the modified form by providing a blank at the end. The fifteen other test items on this sheet were of the modified form, but none contained the correct response among those given, a fact which was intended to be obvious to the pupil. These fifteen items were distributed at random among the other sixty in order to furnish a proper mind-set, since the pupil saw clearly that he must write the proper response in the blank at the end of some of the items on this sheet.

The subjects were 206 pupils in the University of Michigan High School, grouped in six classes: one seventh- and one eighth-grade class in general science, one ninth-grade class in general biology, and one tenth-grade class in advanced biology[1] all enrolled in the regular winter session of 1926-27; and two eighth-grade classes in general science enrolled in the summer session of 1927.

[1] The authors wish to make grateful acknowledgment of the help of Mrs. Grace K. Gelder, of the University High School Department of Biology, in preparing, administering, and scoring the tests in biology.

Both Sheets I and II were distributed to the members of each class during the same examination period, one sheet being given to the pupil after he had completed and handed in the other. In order to equalize practice effects, Sheet I was given first to three classes, and Sheet II first to the other three.

The directions upon Sheet I were as follows:

In the following statements, underline the one of the five words or phrases which alone will make the statement true. Thus, if the statement is this, "This month is February, March, April, May, June," you would underline June, which is the only one of the five words which will make the statement true. After you had marked it the statement would look like this,

"THIS MONTH IS FEBRUARY, MARCH, APRIL, MAY, JUNE."[2]

UNDERLINE ONE WORD OR PHRASE FOR *EVERY* STATEMENT, BUT DO NOT UNDERLINE *MORE* THAN ONE WORD OR PHRASE FOR ANY STATEMENT. Be sure to mark every statement. If you are in doubt regarding which is the correct word or phrase to underline, guess.

The directions upon Sheet II were:

The following statements are of two types. Each contains five words or phrases followed by a blank space. In some of these statements one of the five words or phrases makes the statement correct. Thus:

"THIS MONTH IS FEBRUARY, MARCH, APRIL, MAY, JUNE, ————,"

With statements of this type, you would underline the word which would make the statement correct, and would write nothing in the blank space. When you had finished completing the statement it would look like this:

"THIS MONTH IS FEBRUARY, MARCH, APRIL, MAY, JUNE, ————,"

because you would have underlined June,[3] since June is the only one of the five words which makes the statement true.

In some of the other statements none of the given words or phrases makes the statement true. Thus:

"THIS MONTH IS FEBRUARY, MARCH, APRIL, MAY, JULY, ————."

With statements of this type you would write in the blank space the necessary word, or as many words as are necessary to make the statement correct. In the statement given here you would write June[4] in the blank, thus:

"THIS MONTH IS FEBRUARY, MARCH, APRIL, MAY, JULY, *JUNE*."

IN EACH STATEMENT THAT FOLLOWS DO NOT UNDERLINE ANY WORD OR PHRASE *IF YOU FILL THE BLANK*, BUT BE SURE EITHER TO UNDERLINE A WORD OR PHRASE OR FILL THE BLANK FOR EACH STATEMENT. DO *NOT* FILL THE BLANK UNLESS *NONE* OF THE FIVE WORDS OR PHRASES GIVEN MAKES THE STATEMENT CORRECT.

---

[3] June was changed to August for the summer-school classes.
[4] *Ibid.*
[5] *Ibid.*

It will be noted that the sole difference in the directions for the two test sheets is that Sheet I requires the pupil only to decide which of the given words or phrases makes the statement correct, and then to underline this word or phrase; whereas, with Sheet II, he must not only decide whether the correct response appears in the statement, but must underline it if it is present, and if not, must write it in the blank.

The findings from the investigation were based upon the pupils' responses to the sixty identical items common to their respective Sheets I and II; their responses to the remaining fifteen items on each sheet were disregarded entirely in so far as the investigation was concerned.

FINDINGS

Table I seems to indicate that *the modified form requires only slightly, if any, longer to administer than the multiple-response.* It will be noted that each class required more time in completing whichever form was administered to it first. Thus, with the three classes which were given the multiple-response first, the 12½ percentile pupil of the seventh-grade

TABLE I

TIME IN MINUTES AND SECONDS REQUIRED BY THE 12½ PERCENTILE PUPIL, THE MEDIAN PUPIL, AND 87½ PERCENTILE PUPIL TO FINISH EACH FORM OF THE TEST

CLASSES	MULTIPLE-RESPONSE			MODIFIED FORM		
	12½ Percentile Pupil (Min.-Sec.)	Median Pupil (Min.-Sec.)	87½ Percentile Pupil (Min.-Sec.)	12½ Percentile Pupil (Min.-Sec.)	Median Pupil (Min.-Sec.)	87½ Percentile Pupil (Min.-Sec.)
VII	*18–20**	*21–50*	*30–00*	15–35	18–40	24–20
VIII	*17–00*	*24–30*	*33–00*	15–15	19–50	25–20
IX	*18–20*	*23–50*	*30–00*	16–45	21–15	28–30
X	17–20	19–45	25–20	*24–30*	*33–10*	*37–30*
VIII$_{ss1}$	11–30	15–50	18–30	*19–45*	*27–30*	*36–00*
VIII$_{ss2}$	11–00	13–45	20–30	*22–00*	*28–30*	*43–00*
All classes combined	15–40	21–30	29–25	16–10	21–05	31–45
Weighted averages	16–47	21–24	28–20	17–38	22–35	29–25

*The italics indicate the test-form which was administrated first to each respective class.

finished the multiple-response and the modified form in respectively 18 minutes 20 seconds, and 15 minutes 35 seconds; the median pupil in the eighth-grade finished the multiple-response and the modified form in respectively 24 minutes 30 seconds, and 19 minutes 50 seconds; the 87½ percentile pupil of the ninth-grade finished the multiple response and the modified form in respectively 30 minutes 0 seconds, and 28 minutes 30 seconds. Similar time differences for the 12½ percentile, the median, and the 87½ percentile pupil of each of these three grades are without exception in favor of the modified form. But with the tenth-, the eighth-$_{ss1}$, and the eighth-grade$_{ss2}$, who took the modified form first,

exactly the opposite is true; for these three grades the time differences shown in Table I are all in favor of the multiple-response. No significant differences in time requirements are shown, moreover, when the two hundred six pupils composing all six groups are combined in one time-distribution. An advantage is then found in favor of the multiple-response in the case of the 12½ percentile and the 87½ percentile pupil, and in favor of the modified form in the case of the median pupil. Although somewhat questionable from a statistical standpoint, a computation was made of the weighted averages[5] of the time requirements. The advantages thus found are all in favor of the multiple-response, but they are slight—only about 6 percent for the 12½ percentile pupil, about 5 percent for the median pupil, and about 3½ percent for the 87½ percentile pupil.

Table II presents evidence to show that *the modified form is more difficult for the pupils than the conventional multiple-response.* While

TABLE II

COMPARISON OF THE SCORES FROM TESTS OF THE MULTIPLE-RESPONSE AND THE MODIFIED FORM

	Classes					
	Grade VII	Grade VIII	Grade IX	Grade X	Grade VIII$_{sel}$	Grade VIII$_{unsel}$
Number in Class	51	52	47	24	16	16
Range in Scores						
Multiple-Response	21–55	15–51	19–55	24–56	36–58	22–50
Modified Form	20–52	16–52	21–55	19–50	27–54	19–47
Mean of Scores						
Multiple-Response	36.02	39.10	42.11	37.88	46.69	39.19
Modified Form	33.45	36.46	41.66	33.42	38.75	32.19
S. D. of Scores						
Multiple-Response	± 7.38	± 8.47	± 8.87	± 8.28	± 5.45	± 9.10
Modified Form	± 8.38	± 9.48	± 8.46	±10.38	± 7.00	± 7.91
S. D. of Means						
Multiple-Response	± 1.04	± 1.17	± 1.29	± 1.69	± 1.36	± 2.27
Modified Form	± 1.17	± 1.32	± 1.23	± 2.12	± 1.75	± 1.97
S. D. of S. D.'s						
Multiple-Response	± .73	± .83	± .92	± 1.20	± .96	± 1.61
Modified Form	± .83	± .93	± .87	± 1.50	± 1.24	± 1.40
M. Mult.-Response — M. Modified	+ 1.65	+ 1.50	+ .25	+ 1.65	+ 3.58	+ 2.33
S. D. D. of Means						
S. D. Mult.-Response — S. D. Modified	−.90	−.81	+.32	−1.09	−.99	+.56
S. D. D. of S. D.'s						

[5] The weighted averages were computed thus: The median time on the multiple-response, of, e. g., the 12½ percentile pupil of each grade, was multiplied by the number of pupils in that grade. The sum of the six products thus obtained was divided by the total number of pupils, 206, to give the weighted-average median time. Weighted averages were obtained in this way for the 12½ percentile, the median, and the 87½ percentile pupil on both forms.

the ranges in scores for all six classes were approximately equal with both forms, scored as right and wrong, for example, 15 to 51 points for the multiple-response and 16 to 52 points for the modified form with the eighth-grade; 22 to 50 points for the multiple-response and 19 to 47 points for the modified form with the eighth-grade$_{ss_2}$, etc., the mean scores were in every class lower for the modified form than for the conventional multiple-response. Thus for the seventh grade, the mean on the multiple-response was 36.02 points and for the modified form, 33.45 points; for the tenth grade, 37.88 points in the multiple-response and 33.42 points on the modified form, etc. When, moreover, the difference of the mean scores is divided by the standard deviation of the difference of the means,[6], [7] the results indicate the degree of significance of the greater difficulty of the modified form: Thus, with the eighth-$_{ss_1}$, the eighth-$_{ss_2}$, the seventh-, the eighth-, and the tenth-grades, the chances are respectively more than 999, about 990, 950, 933, and 950 out of a thousand that if the investigation were repeated the mean scores with the modified form would again be lower than with the multiple-response. Here, then, the greater difficulty of the modified form may be considered "practically certain" with the eighth-grade$_{ss_1}$, "probable" with the eighth-grade$_{ss_2}$, and "possible" with the eighth-, the seventh-, and the tenth-grades.

Table II offers evidence which seems to indicate that the *modified form differentiates between the good and the poor pupils at least as well as the multiple-response.* When the difference of the standard deviations is divided by the standard deviation of the differences of the standard deviations,[8] the results show slight values in favor of the modified form with four of the groups, and two smaller values in favor of the multiple-response. Thus, with the seventh-, the eighth-, the tenth-, and the eighth-grade$_{ss_1}$, the chances are respectively about 816, 791, 862, and 839 out of a thousand, and with the ninth-, and the eighth-grade$_{ss_2}$ respectively about 375 and 288 out of a thousand that if the investigation were re-

---

[6] $S.D._M = \sqrt{\left(\frac{S.D._1}{\sqrt{N}}\right)^2 + \left(\frac{S.D._2}{\sqrt{N}}\right)^2}$

[7] For convenience, this partial formula is here used; the complete formula would merely tend to increase, somewhat, and therefore to make somewhat more significant the results obtained from this partial formula.

[8] $S.D._{S.D.} = \sqrt{\left(\frac{S.D._1}{\sqrt{2N}}\right)^2 + \left(\frac{S.D._2}{\sqrt{2N}}\right)^2}$ See also footnote 7.

peated the differences would again be in favor of the modified form. Only the results with the tenth-grade, however, can here be considered as "possible."

Further evidence that the modified form differentiates between the good and the poor pupils at least as well as the multiple-response is indicated in Table II by the greater standard deviations of four of the six classes with the modified form than with the multiple-response.

Table III furnishes a crude estimate of validity which will later be checked by means of a more detailed study. An inspection of the mean scores made on both forms by certain groupings of pupils in each grade, selected on the basis of teacher-judgment without reference to the scores made on these tests, indicates that *both forms of test are valid,* since each form differentiates *consistently* between the best, the medium, and the poorest pupils in each class. Thus, in the seventh-grade, on the multiple-response the best 20 percent of the pupils averaged 8.31 points higher

TABLE III

COMPARISON OF THE MEAN SCORES IN THE MULTIPLE-RESPONSE TEST AND THE MODIFIED FORM, OF VARIOUS GROUPS OF PUPILS SELECTED ON THE BASIS OF TEACHERS' JUDGMENTS WITHOUT REGARD TO TEST SCORES

CLASSES	MEAN SCORES OF BEST 20 PERCENT OF PUPILS	MEAN SCORES OF MIDDLE 60 PERCENT OF PUPILS	MEAN SCORES OF POOREST 20 PERCENT OF PUPILS
Grade VII			
Multiple-Response	43.90	35.59	29.70
Modified Form	41.80	32.16	28.10
Difference	2.10	3.43	1.60
Grade VIII			
Multiple-Response	48.00	40.00	27.30
Modified Form	47.70	36.38	25.50
Difference	.30	3.62	1.80
Grade IX			
Multiple-Response	50.44	42.97	31.00
Modified Form	50.22	42.41	30.67
Difference	.22	.56	.33
Grade X			
Multiple-Response	48.00	37.57	28.00
Modified Form	45.00	33.29	22.20
Difference	3.00	4.28	5.80
Grade VIII$_{\text{sel}}$			
Multiple-Response	54.33	46.00	39.33
Modified Form	48.00	38.00	32.00
Difference	6.33	8.00	7.33
Grade VIII$_{\text{sel}}$			
Multiple-Response	50.67	37.10	34.67
Modified Form	43.00	30.10	28.33
Difference	7.67	7.00	6.34

than the middle 60 percent of the pupils, and the middle 60 percent of the pupils, 5.89 points higher than the poorest 20 percent of the pupils;

on the modified form, the best averaged 9.64 points higher than the middle pupils, and the latter, 4.06 points higher than the poorest pupils.

The data in Table III also seem to indicate that *the modified form is relatively at least as valid as the multiple-response*, since it differentiates between the best, the medium, and the poor pupils at least as well as the multiple-response. With the eighth-grade$_{ss2}$, the best 20 percent of the pupils, selected on the basis of teacher-judgment without reference to scores on these tests, made a mean score of 7.67 points lower on the modified form than on the multiple-response, the middle 60 percent made a mean score of 7.00 points lower, and the poorest 20 percent a mean score 6.34 points lower. If with all six classes there had been a similar, progressively smaller difference between the mean scores of these three groups on the two tests, the evidence would indicate a consistent superiority of the multiple-response over the modified form as a means of differentiating between the good and the poor pupils. But the data fail to reveal any such consistent superiority of the multiple-response; in fact the results from the tenth-grade indicate the contrary, since the corresponding differences are progressively *larger* instead of smaller. Here the differences in the mean scores are for the best 20 percent, 3.00 points; for the middle 60 percent, 4.28 points; and for the lowest 20 percent, 5.80 points. The remaining data in Table III are no less favorable to the modified form than to the multiple-response. The differences in means on both forms, between the scores of the best 20 percent and the middle 60 percent favor the modified form in the seventh-, eighth-, ninth-, and eighth-grade$_{ss1}$, since they are progressively larger; while with these same classes the differences in means on both forms between the scores of the middle 60 percent and the poorest 20 percent favor the multiple-response, since they are progressively smaller.

The modified form as tested *is equally as reliable as the conventional multiple-response*. Table IV shows a considerable variation in the re-

TABLE IV

COMPARISON OF THE RELIABILITY COEFFICIENTS OF THE TWO HALVES RESPECTIVELY OF THE MULTIPLE-RESPONSE TEST AND THE MODIFIED FORM, COMPUTED BY THE PEARSON PRODUCT-MOMENT FORMULA

	Classes					
	Grade VII	Grade VIII	Grade IX	Grade X	Grade VIII$_{ss1}$	Grade VIII$_{ss2}$
Number in Class	51	52	47	24	16	16
Reliability r						
Multiple-Response	.35	.78	.78	.97	.32	.84
Modified Form	.57	.68	.87	.93	.49	.72

liability coefficients of the two forms when correlations are computed between the scores on the two halves of the tests. When the reliability coefficient is computed for all six classes combined as one group, however, the coefficient is found to be .70 for the multiple-response and .73 for the modified form, a difference of .03 in favor of the modified form. But the relatively small number of cases and classes used in this investigation, and the size of the probable error as compared with the difference, prevent definite conclusions regarding the relative reliabilities of the two forms.

Table V indicates that *the reliability of the modified form* as computed by Brown's formula from these tests in science, *is nearly as high*

TABLE V

COMPARISON OF THE RELIABILITIES OF ONE HUNDRED ITEMS OF VARIOUS TYPES OF TESTS REPORTED BY RUCH, AND BY MCCLUSKY AND CURTIS, WITH RELIABILITIES OF ONE HUNDRED ITEMS OF THE AUTHORS' MULTIPLE-RESPONSE TEST AND MODIFIED FORM TEST, COMPUTED BY BROWN'S FORMULA*

TYPE AND TEST	NUMBER OF PUPILS	RELIABILITY r, HALF VERSUS HALF†	RELIABILITY r, 100 ITEMS
Ruch's			
Recall	562	.81 ± .010	.90
Five-response	137	.80 ± .020	.89
Three-response	134	.60 ± .037	.75
Two-response	135	.74 ± .027	.85
True-False	133	.56 ± .040	.71
McClusky-Curtis			
True-false (science)	115	.53 ± .045	.764‡
Modified Form (science)	115	.76 ± .026	.904‡
Authors'			
Five-response	206	.70 ± .024	.859‡
Modified Form	206	.73 ± .022	.851‡

$$*r = \frac{nr}{1 + (n - 1)\,r}$$

†Fifty items for each of Ruch's tests; twenty-five items each for McClusky and Curtis' tests; and sixty items each for the authors' tests.

‡Corrected for dispersion, and thus made comparable to these r's reported by Ruch, by means of the formula $\frac{r_{\infty}}{\Sigma_{\infty}} = \frac{\sqrt{r_{11}(1 - R_{11})}}{\sqrt{R_{11}(1 - r_{11})}}$ Kelley, Truman L.: *Statistical Method* (New York: The Macmillan Co., 1923), p. 221.

as the reliability of the recall and multiple-response tests experimented upon by Ruch,[9] *and as the modification of the true-false* experimented upon by McClusky and Curtis.[10] The reliabilities of Ruch's recall and

[9] Ruch, Giles M. *The Improvement of the Written Examination* (New York: Scott Foresman & Co., 1925), p. 111.

[10] McClusky, Howard Y. and Curtis, Francis D. "A Modified Form of the True-False Test," *Journal of Educational Research*, XIV (1926), 213-224.

five-response tests were respectively .90 and .89; of the modified form of the true-false (science) .90; of the authors' five-response tests, .86, and of the modified form of the multiple-response, .85.

Several other statements seem justified by the authors' extensive use of both these forms, although no objective evidence supporting them can here be presented. It apparently takes no longer to construct test items of the modified form than of the multiple-response; also, pupils are able to score this form readily, though they, as well as the examiner, require a little more time to score the modified form than the multiple-response. The modified form seems, moreover, to develop in the pupils to a greater extent than the conventional type, the habit of critical scrutiny and analysis of every item; it tends, furthermore, to eliminate whatever elements of chance may lurk in the multiple-response.

An examination of individual test papers indicates, too, that the modified form is superior to the multiple-response for diagnosis of individual and class difficulties and weaknesses.

SUMMARY

In so far as the results of this investigation may be indicative, it seems reasonable to conclude that this modified form of the multiple-response test is more difficult than the conventional multiple-response. It is inferior in that it requires somewhat more time to correct and score, though it requires very little, if any, longer than the conventional form to administer. It is at least as good in the following respects: (1) It takes no longer to construct. (2) It differentiates at least as well between the good and the poor pupils. (3) It is probably as valid. (4) It is at least as reliable as the conventional form. It is superior to the conventional form in the following respects: (1) It possesses greater value for diagnosis of individual and class difficulties and weaknesses. (2) It stimulates more thinking on the part of the pupil while he is taking the test. (3) It tends to eliminate whatever guessing factor may operate in the multiple-response.

UNFAMILIAR SITUATIONS AS A DIFFICULTY IN SOLVING ARITHMETIC PROBLEMS

CARLETON WASHBURNE
Superintendent of Schools
and
MABEL VOGEL MORPHETT
Director of Research Department, Winnetka, Illinois

CAN children solve arithmetic problems presenting situations within their own experience more readily than they can solve problems involving unfamiliar situations? In other words, all other elements being equal, is the *situation* a real factor in the difficulty of an arithmetic problem?[1]

To answer this question a problem solving test made up of eight pairs of problems was devised, at the fifth-grade level. The two problems that made up a pair involved the same arithmetic processes, numbers of the same order of difficulty, and the same sentence structure and length; they differed only in the situation involved (see Table I). Six of the pairs consisted of two-step problems, i.e., problems requiring two different arithmetic processes; the remaining two pairs required only one step for their solution. The arithmetic processes involved were: column addition; subtraction with carrying; simple multiplication; short division; compound multiplication; and long division.

The two members of each pair were separated and the sixteen problems mixed and presented in random order.

The test was mimeographed and given to 441 children in the fifth grades of the public schools of Wilmette, Batavia, Dolton, Highland Park, and Winnetka (all in Illinois) and of the Francis W. Parker School (Chicago).

The tests were scored on the basis of process and accuracy. The process was counted correct if the child showed that he knew how to work the problem, but had made some error in calculation. A child had to have the correct answer in order to be marked correct in accuracy.

Table I gives the results of the test, for both process and accuracy. Comparable problems are shown, together with the percent correct in process and accuracy for both the familiar and unfamiliar situations.

Let us take the first two examples on the table. The processes used are column addition and simple multiplication. The wording of the two problems is nearly identical. The number difficulty of the two problems

[1]This study was carried out by the Winnetka Research Department to supplement the studies of the Committee of Seven reported by Carleton Washburne and Raymond Osborne in the *Elementary School Journal* for November, 1926.

TABLE I

PROBLEM SOLVING TEST

FIFTH GRADE

PAIR 1

UNFAMILIAR ADDITION = MULTIPLICATION

A steam shovel made an excavation for a building. It dug out 26 truck-loads on Monday, 43 on Tuesday, and 56 on Wednesday. Each truck-load weighed 3 tons. The owner of the shovel is going to charge his employers for the work. How many tons will he have to charge to them?

FAMILIAR ADDITION = MULTIPLICATION

Some of the children are going to make booklets for a bazaar. Miss Graham's room is going to make 35 booklets; Miss Leap's room, 42 booklets; and Miss Damberg's, 46 booklets. There will be 4 sheets in each booklet. John is going to buy the paper for all the rooms. How many sheets will he have to buy for them?

PERCENT CORRECT

Process		Accuracy
Unfamiliar	51	45
Familiar	67	60

PAIR 2

UNFAMILIAR ADDITION = DIVISION

In August a farmer used 64 bushels of prepared feed; in September, 72 bushels; in October, 73 bushels; and in November, 63. He has 8 cows. How much prepared feed did each cow eat?

FAMILIAR ADDITION = DIVISION

On Monday we took in 66 pins at our circus; on Tuesday, 96 pins; on Wednesday, 73; and on Thursday, 61. There are 8 of us in the circus. How many pins does each of us get?

PERCENT CORRECT

Process		Accuracy
Unfamiliar	73	58
Familiar	75	60

PAIR 3

UNFAMILIAR MULTIPLICATION = SUBTRACTION

A contractor needed 820 bolts in order to complete a building job. He has 25 bolts in each of 30 packages. How many more bolts does he need to complete the job?

FAMILIAR MULTIPLICATION = SUBTRACTION

Jack needed to sell 920 copies of Liberties in order to win a baseball glove. He sold 44 Liberties each week for 20 weeks. How many more copies does he have to sell to win the glove?

PERCENT CORRECT

Process		Accuracy
Unfamiliar	66	61
Familiar	65	62

PAIR 4

UNFAMILIAR ADDITION = SUBTRACTION

A telephone company had 62 spools of wire. Yesterday it received 11 from a nearby factory. Today it received 39 more from there. Then 15 of them were used. How many spools of wire should the company have now?

FAMILIAR ADDITION = SUBTRACTION

I had 55 marbles. Yesterday my daddy gave me 23 new ones. Today my brother gave me 35 of his. Then I gave my chum 15 of them. I wonder if I have lost any. How many marbles should I have now?

PERCENT CORRECT

Process		Accuracy
Unfamiliar	74	63
Familiar	85	71

PAIR 5

UNFAMILIAR MULTIPLICATION

A merchant sold 20 bags of charcoal. Each bag held 35 pieces. How many pieces of charcoal did he sell?

FAMILIAR MULTIPLICATION

The girls have to make 30 boxes of taffy. Each of the boxes holds 25 pieces. How many pieces of taffy do they have to make?

PERCENT CORRECT

Process		Accuracy
Unfamiliar	91	79
Familiar	88	79

PAIR 6

UNFAMILIAR DIVISION=MULTIPLICATION

A public typist wrote some letters for a lawyer. She typed a letter of 357 words in 7 minutes. She wants to know how long a letter she could write if she typed for 9 minutes.

FAMILIAR DIVISION=MULTIPLICATION

John and his father took a trip to Canada. One day they drove 147 miles in 7 hours. John wants to know how far they could go if they drove 9 hours the next day.

PERCENT CORRECT

Process		Accuracy
Unfamiliar	44	41
Familiar	43	40

PAIR 7

UNFAMILIAR DIVISION

19 chemists bought 627 Florentine flasks. Each chemist is to have the same number of flasks. How many flasks should each chemist have?

FAMILIAR DIVISION

18 Boy Scouts caught 396 fish while on a fishing trip. Each scout is to have the same number of fish. How many fish should each scout take home?

PERCENT CORRECT

Process		Accuracy
Unfamiliar	82	65
Familiar	91	81

PAIR 8

UNFAMILIAR SUBTRACTION=DIVISION

A coke plant had to turn out 673 tons of coke for a steel company. It has already produced 129 tons. There are 32 ovens in the plant. How many more tons will each oven have to produce?

FAMILIAR SUBTRACTION=DIVISION

My committee has been asked to dress 363 pencils for the bazaar. We have already dressed 149 pencils. There are 23 girls on the committee. How many more pencils will each girl have to dress?

PERCENT CORRECT

Process		Accuracy
Unfamiliar	53	45
Familiar	71	54

is exactly comparable. The only variable that seems to affect the solution is the familiarity of the situation. The unfamiliar problem involving truck loads of dirt for a building excavation was correctly attacked by 51 percent of the children and was correctly solved by 45 percent of the children. The familiar problem involving a more childlike situation

in a classroom was correctly attacked by 67 percent of the children and was correctly solved by 60 percent of the children. The problem involving a familiar, childlike situation is attacked and solved correctly by a larger percentage of children.

In pairs 1, 4, 7, and 8 a decidedly larger percentage of children was able to solve the problems containing familiar situations than were able to solve problems containing unfamiliar situations. In pairs 2, 3, 5, and 6, there is no significant difference between the scores of the two sets of problems.

In three cases the failure to get a significant difference may lie with the test. Pair 2, for example, compares feeding cows with collecting pins for a circus. The former, intended to be unfamiliar, is probably fully as familiar as the latter. The difference between selling bags of charcoal and filling bags with taffy (pair 5) is probably not significant. Typing letters (pair 6) seems *a priori* less familiar than taking auto rides, but may not really be. But certainly pair 3 should present a contrast—selling Liberties must be more familiar than completing a building with bolts; it is difficult to understand why both of these situations caused the same amount of difficulty.

Notwithstanding these exceptions, the fact remains, however, that whenever a significant difference occurs in the percentage of children solving correctly the two sets of problems, it is always the problem involving the more familiar, childlike situation which gets the higher score.

Table II summarizes the data. This table shows that where one form or the other of a problem is missed, there is just about twice as good a chance (429 to 218) of the child missing the form involving an unfamiliar situation as there is of his missing the one where the situation is familiar.

TABLE II

THE NUMBER OF RIGHTS AND WRONGS FOR EACH PAIR OF PROBLEMS

PROBLEM	UNFAMILIAR RIGHT FAMILIAR RIGHT	UNFAMILIAR WRONG FAMILIAR WRONG	UNFAMILIAR RIGHT FAMILIAR WRONG	UNFAMILIAR WRONG FAMILIAR RIGHT	TOTAL
1	197	109	35	100	441
2	289	79	32	41	441
3	247	105	48	41	441
4	298	40	27	76	441
5	369	23	31	18	441
6	181	232	16	12	441
7	351	31	9	50	441
8	221	109	20	91	441
Totals	2153	728	218	429	3528

The data presented herein bear out the common sense conclusion, and the preliminary experiments of the Committee of Seven, in showing that part of the difficulty that children have in applying their arithmetic to textbook problems lies in the problem, not the children. Textbook makers and teachers alike should make the problems they expect children to solve, childlike and real. The three following criteria, all of which were applied in the making of the "familiar" problems of this experiment, should prove helpful in problem making:

1. The situation should be familiar and childlike—i.e., one in which the child can easily imagine himself as being.

2. The principal character in each problem should have a reason for solving it—i.e., if the child were in the situation described in the problem, he would feel a real need for finding the solution for which the problem calls.

3. The vocabulary and sentence structure of the problem should, of course, be within the reading ability of the child who is expected to solve it.

A TABLE FOR PREDICTING THE VALIDITY AND RELIABILITY COEFFICIENTS OF A TEST WHEN LENGTHENED

HAROLD A. EDGERTON and HERBERT A. TOOPS
Ohio State University

KEEPING constant the factors of difficulty and type of content, an increase in the length of a test may be expected to raise both its validity and reliability coefficients.

In test construction, for example, the question often arises, "Shall I increase the length of the present test in order to increase the validity and reliability of the predictions, or shall I try a different test, that is, one made up of a different content?" The accompanying table will help to answer this problem. For, with a knowledge of the validity and reliability coefficients to be expected in a lengthened test of feasible length, one may adopt an arbitrary rule, such as the following, to guide his procedure:—"If the addition of one minute of test time will not add .005 to the reliability coefficient or .002 to the validity coefficient, then it is not worth while to increase the length of the present test further, and one should search then instead for different test content."[1]

[1] The new items, when added, should parallel the present items in difficulty and should be "sandwiched in" at the proper difficulty levels. If t is the increase in the time limit of the present test, the increase per minute in either validity or reliability is $1/t$ times the actual increase in validity or reliability respectively. If a three-minute test has an initial reliability of .60, then doubling the test leads us to expect a reliability of the lengthened test of .750; tripling, .818; quadrupling, .857; quintupling, .882; sextupling, .900; and septupling, .913. Since the increase at each step in time is 3 minutes, $t = 3$, and it follows that $1/t = \frac{1}{3}$. One-third of (.750 — .60) is more than .005; so also is $\frac{1}{3}$ (.818 — .750); and $\frac{1}{3}$ (.857 — .818); $\frac{1}{3}$ (.882 — .857); and $\frac{1}{3}$ (.900 — .882); but not $\frac{1}{3}$ (.913 — .900) which is only .004.

Therefore, as far as reliability is concerned, we may say that it is not particularly profitable to make the lengthened test contain more than six times the present number of items. If the test originally contained 50 items, the lengthened test would now contain 300 items. For each item of the original test we need to supply five other items of the same difficulty and kind of test content, or 250 more questions. The number of multiples, n, evidently refers then to an increase in the time limit as well as to an increase in the number of questions. If the original test rapidly becomes so difficult that no test subject attempted the last five items, say, it may suffice merely to add five test items of parallel difficulty to each of the first forty-five questions and to sextuple the time limit on the test. Time required in other operations than working on the test content,—time for reading directions, for example—should be excluded from t in the above computations involving the expression, $1/t$. The final test, then, would consist of 300 items; would require eighteen minutes; and would have a reliability of .90. In general it would cost six times as much to print such a test; six times as much to score it, and so on in like proportion. This cost consideration might lead us to abandon the test even though the reliability is now acceptable. The more pertinent consideration is always validity rather than reliability, for no test, however reliable, if used by itself, is of any worth unless its validity is acceptable.

The accompanying table will show by inspection all the various interrelationships of length of test, validity coefficients and reliability coefficients between certain wide limits. The table gives the expected validity and reliability coefficients, as dependent upon (a) the present reliability, r_{AB}, and upon (b) the increased number of questions for lengthened tests which would contain twice the present number of questions ($n = 2$), three times the present number of questions ($n = 3$), up to and inclusive of a test which would contain fifteen times the present number of questions ($n = 15$). This more than covers all commonly needed cases.

USE OF THE TABLE

Suppose we have a three-minute test which has a reliability coefficient (r_{AB}) of .60, and which correlates .40 with a criterion of, say, scholarship. The validity coefficient upon lengthening is given by the following formula, (A) : $-r_{I\,(nA)} = K.r_{IA}$. In which, $r_{I\,(nA)}$ is the validity coefficient of the lengthened test; K is the plain-faced number of the table, corresponding to the given initial reliability coefficient, r_{AB}, and the number of multiples, n, contained in the longer test; r_{IA} is the initial validity coefficient of the test. In the above example, where, $r_{IA} = .40$, and, $r_{AB} = .60$ by following horizontally across the page in the row, $r_{AB} = .60$, we may note the following successive values of K (plain-faced numbers, always larger than 1.00) as dependent upon n:—When $r_{AB} = .60$, and $n = 2$, $K = 1.118$; $n = 3$, $K = 1.168$; $n = 4$, $K = 1.195$; $n = 5$, $K = 1.213$; etc. This means, by formula (A) above, that if we double the length of the test, we may expect the thus lengthened test to correlate with scholarship to the extent of 1.118 times the present amount, i.e., 1.118 (.40) = .45; with the test tripled in length, 1.168 (.40) = .47; with the test quadrupled in length, 1.195 (.40) = .48; and with the test quintupled in length, 1.213 (.40) = .485, etc., etc. The above series shows that little additional validity is to be gained after this test has been tripled or quadrupled in length. If the test were tripled in length, number of questions, and time limit, so that it were then a nine-minute test, the test would have a validity coefficient of .47; and the reliability of this nine-minute test would be .818.

This reliability coefficient is an italicized entry found in the table in the 60-row and in the column, $n = 3$ *just beneath* the above found K-entry of 1.168. The value of the increased reliability coefficient is thus found directly and without any computation whatsoever. These italicized figures are always less than 1.00.

We may be interested to know, in the above example, to what extent we may have to increase the length of the test in order to secure a reliability coefficient of, say, .90. Tracing the reliability coefficients (italicized figures) in the $r_{AB} = .60$ row, to the right until we find the first entry as large as .90, we find the entry .900, and directly beneath it at the bottom edge of the chart (or at the head of the same column) the desired answer, $n = 6$. If our desired reliability is .90, we shall have to make the three-minute test into an eighteen-minute test. If we wish to have a reliability coefficient of .95, then the test must be made to contain thirteen times the present number of items.

We may also have in mind a desired, or "ideal" or minimally acceptable validity coefficient. If the above results as to reliability and validity were obtained in a thirty-minute university intelligence test, we might not feel satisfied with a validity less than .50, perhaps. Substituting the desired validity coefficient .50, for $r_{I\,(nA)}$, and .40 for r_{IA}, in formula (A), we obtain, $.50 = K(.40)$. Solving for K, we obtain, $K = 1.25$. If, now, we trace across to the right in the bold-faced numbers of row .60, we find the entry, 1.250 corresponds exactly to $n = 10$. That is, this half-hour university intelligence test would need to be increased to 10 times 30 or 300 minutes, or 5 hours, in length in order to obtain a validity coefficient of .50. Evidently, the improvement is not worth the cost in time. Although the five hour test has a high reliability (.938), the validity is still low. Five hours of test time in this case can be far better spent on different content.

The validity coefficient improves much more slowly than the reliability coefficient as the result of lengthening the test. By looking in the column headed ∞ we note that if the above half-hour university intelligence test were made infinitely long it would have a validity only 1.291 times the present validity coefficient, or $1.291\ (.40) = .5164$. If five hours of testing give us a validity of .50 and if an infinite number of hours of testing give us .516, then obviously it is not worth the while, so far as validity is concerned to make the test much longer than five hours in length. The value of the reliability coefficient for any test when made infinitely long is always 1.00. The definition of a "true score" is thus determined; a true score is any score which is perfectly reliable, or it is the average score made on an infinite number of repetitions of a test.

THE FORMULAS

Lack of space prohibits us from deriving the equations used. Formula (B) for the increased reliability coefficient is:—

$$r_{(nA)(nA)} = \frac{n \cdot r_{AB}}{1 + (n-1)\, r_{AB}}$$

The formula for the increased validity coefficient is adequately defined by substituting in Formula (A) the following formula (C) for K:

$$K = \frac{n}{\sqrt{n + n(n-1)\, r_{AB}}} = \sqrt{\frac{n}{1 + (n-1)\, r_{AB}}}$$

It is desirable to keep in mind the assumptions involved in the development of Formulas (A) and (B). Formula (A), or (C), assumes:

1. That the correlation of the first form of the test with the criterion, (r_{1A}) is of the same numerical value as the *average* correlation of all other multiples with the same criterion. This is a fair assumption probably, at least for the length of test time involved for most test scales,—a length of time probably ordinarily quite too short for practice effects to play too great a part. Inasmuch as one ordinarily cannot obtain a reliability coefficient without having two alternative forms of the test prepared, thus making it possible to correlate both forms of the test with the criterion, it is better to take for r_{1A} of formula (A) the *average of the two validity coefficients;* and if more than two are available, the average of all of them.
2. That the standard deviations of the successive forms, or multiples of the original tests are all equal. If the new questions are equally as well scaled as the original, then this is also probably a fair assumption.
3. That the correlation, r_{AB}, the reliability coefficient obtained from the first two forms of the test, is of the same numerical value as the *average* of all such possible intercorrelation reliability coefficients. If practice does not affect the results within the test time limit of the total test, then this is a fair assumption. As in assumption 1 above, one should use for r_{AB} in both formulas (A) and (B) the *average of any and all reliability coefficients of the different forms he has at hand, or can readily obtain.* It can readily be seen that, with n alternative forms of a test given, one of our pressing needs is a formula for telling us what is the average Pearson intercorrelation coefficient of the n forms, without solving all the $\frac{n(n-1)}{2}$ intercorrelation coefficients. This formula is now available.[2]

[2] Edgerton, H. A. and Toops, H. A. "A Formula for Finding the Average Intercorrelation Coefficient of Unranked Scores Without Solving Any of the Individual Intercorrelations." *Journal of Educational Psychology.* January, 1928.

In formula (B) above, assumptions 2 and 3 are both made, and no others.

It will be noted that, the *lower* the initial reliability of a test, the greater is the hope of improving it, both in reliability and validity, by increasing the number of test elements. We also see that a test of low initial validity is seldom worth attempting to salvage by increasing its length.

The limits of improvement in validity are found by taking limits, as n approaches infinity, in formula (C). This yields formula (D): Maximum possible validity $= r_I(\infty x) = \frac{r_{IA}}{\sqrt{r_{AB}}}$. This is to be defined as the correlation of a criterion and a true test score, or the correlation of an infinite number of forms of the test with the criterion. Formula (D) yields the K-values of the column, $(n = \infty)$, of the table.

Taking limits similarly in formula (B) we obtain $r_{(\infty X)(\infty X)} = 1.00$ as the result. True test scores are thus perfectly reliable; they yield a perfect measure of the test subject's ability on that test, but do not, of course, correlate perfectly with any criterion. Thus it follows that we may have a test of perfect reliability ($r_{AB} = 1.00$) but of zero validity ($r_{IA} = 0$); the reverse however, is never possible. The number of hairs on one's head counted on successive days, might be expected to have about perfect reliability but when used as a test of scholarship, might be expected not to correlate with scholarship at all!

CONCLUSIONS

A table has been constructed which will enable one to determine by inspection, and with no computation beyond one simple multiplication, the increased validity and reliability of a test when lengthened to any desired predetermined limit, so long as the test is not lengthened to more than fifteen times its present length. By reversing the formulas, it is readily possible to determine how many multiples of the present length of test are required respectively to secure any predetermined reliability coefficient, say, .80 or .85 or .90 or any predetermined, obtainable, validity coefficient. The limits of improvability both as to validity and reliability are also shown. The table may be used not only with tests, but is also useful in experimental work generally to determine the number of repetitions necessary to produce a desired validity or reliability. Thus, if two examinations (medical, physical, civil service examinations, school examinations, or what not) yield marks which correlate with each

r/ab	2	3	4	5	6	7	8	9	10	11	12	13	14	15	∞	r/ab
.01	1.407	1.714	1.971	2.193	2.390	2.570	2.734	2.887	3.029	3.162	3.288	3.392	3.520	3.627	10.000	.01
	.020	*.029*	*.039*	*.048*	*.057*	*.066*	*.075*	*.083*	*.092*	*.100*	*.108*	*.116*	*.124*	*.132*	*1.000*	
.02	1.400	1.698	1.943	2.152	2.335	2.500	2.649	2.785	2.911	3.028	3.136	3.238	3.333	3.423	7.071	.02
	.039	*.058*	*.075*	*.093*	*.109*	*.125*	*.140*	*.155*	*.169*	*.183*	*.197*	*.210*	*.222*	*.234*	*1.000*	
.03	1.393	1.682	1.916	2.113	2.284	2.436	2.571	2.694	2.806	2.909	3.004	3.092	3.174	3.250	5.774	.03
	.058	*.085*	*.110*	*.134*	*.157*	*.178*	*.198*	*.218*	*.236*	*.254*	*.271*	*.287*	*.302*	*.317*	*1.000*	
.04	1.387	1.667	1.890	2.076	2.236	2.376	2.500	2.611	2.712	2.803	2.887	2.964	3.035	3.101	5.000	.04
	.077	*.111*	*.143*	*.172*	*.200*	*.226*	*.250*	*.273*	*.294*	*.314*	*.333*	*.351*	*.368*	*.385*	*1.000*	
.05	1.380	1.651	1.865	2.041	2.191	2.320	2.434	2.535	2.626	2.708	2.782	2.850	2.913	2.970	4.472	.05
	.095	*.136*	*.174*	*.208*	*.240*	*.269*	*.296*	*.321*	*.345*	*.367*	*.387*	*.406*	*.424*	*.441*	*1.000*	
.06	1.374	1.637	1.841	2.008	2.148	2.269	2.374	2.466	2.548	2.622	2.689	2.749	2.804	2.855	4.082	.06
	.113	*.161*	*.203*	*.242*	*.277*	*.309*	*.338*	*.365*	*.390*	*.412*	*.434*	*.453*	*.472*	*.489*	*1.000*	
.07	1.367	1.622	1.818	1.976	2.108	2.220	2.317	2.402	2.477	2.544	2.604	2.658	2.707	2.752	3.780	.07
	.131	*.184*	*.231*	*.273*	*.311*	*.345*	*.376*	*.404*	*.429*	*.453*	*.475*	*.495*	*.513*	*.530*	*1.000*	
.08	1.361	1.608	1.796	1.946	2.070	2.175	2.265	2.343	2.411	2.472	2.526	2.575	2.620	2.660	3.536	.08
	.148	*.207*	*.258*	*.303*	*.343*	*.378*	*.410*	*.439*	*.465*	*.489*	*.511*	*.531*	*.549*	*.566*	*1.000*	
.09	1.355	1.594	1.775	1.917	2.034	2.132	2.215	2.287	2.351	2.406	2.456	2.500	2.540	2.576	3.333	.09
	.165	*.229*	*.283*	*.331*	*.372*	*.409*	*.442*	*.471*	*.497*	*.521*	*.543*	*.562*	*.581*	*.597*	*1.000*	
.10	1.348	1.581	1.754	1.890	2.000	2.092	2.169	2.236	2.294	2.345	2.390	2.431	2.467	2.500	3.162	.10
	.182	*.250*	*.308*	*.357*	*.400*	*.438*	*.471*	*.500*	*.526*	*.550*	*.571*	*.591*	*.609*	*.625*	*1.000*	
.11	1.342	1.568	1.734	1.863	1.967	2.054	2.126	2.188	2.242	2.289	2.330	2.367	2.400	2.430	3.015	.11
	.198	*.270*	*.331*	*.382*	*.426*	*.464*	*.497*	*.527*	*.553*	*.576*	*.597*	*.616*	*.634*	*.650*	*1.000*	
.12	1.336	1.555	1.715	1.838	1.936	2.017	2.085	2.143	2.193	2.236	2.274	2.308	2.339	2.366	2.887	.12
	.214	*.290*	*.353*	*.406*	*.450*	*.488*	*.522*	*.551*	*.577*	*.600*	*.621*	*.639*	*.656*	*.672*	*1.000*	
.13	1.330	1.543	1.696	1.814	1.907	1.983	2.047	2.100	2.147	2.187	2.222	2.253	2.281	2.306	2.774	.13
	.230	*.310*	*.374*	*.428*	*.473*	*.511*	*.545*	*.574*	*.599*	*.622*	*.642*	*.660*	*.677*	*.691*	*1.000*	
.14	1.325	1.531	1.678	1.790	1.879	1.950	2.010	2.060	2.104	2.141	2.174	2.202	2.228	2.251	2.673	.14
	.246	*.338*	*.394*	*.449*	*.494*	*.533*	*.566*	*.594*	*.619*	*.642*	*.661*	*.679*	*.695*	*.709*	*1.000*	
.15	1.319	1.519	1.661	1.768	1.852	1.919	1.975	2.023	2.063	2.098	2.128	2.155	2.178	2.200	2.582	.15
	.261	*.346*	*.414*	*.469*	*.514*	*.553*	*.586*	*.614*	*.638*	*.660*	*.679*	*.696*	*.712*	*.726*	*1.000*	
.16	1.313	1.508	1.644	1.746	1.826	1.890	1.943	1.987	2.024	2.057	2.085	2.110	2.132	2.152	2.500	.16
	.276	*.364*	*.432*	*.488*	*.533*	*.571*	*.604*	*.632*	*.656*	*.677*	*.696*	*.712*	*.727*	*.741*	*1.000*	
.17	1.307	1.496	1.628	1.725	1.801	1.862	1.911	1.953	1.988	2.018	2.045	2.068	2.088	2.107	2.425	.17
	.291	*.381*	*.450*	*.506*	*.551*	*.589*	*.621*	*.648*	*.672*	*.693*	*.711*	*.727*	*.741*	*.754*	*1.000*	
.18	1.302	1.485	1.612	1.705	1.777	1.834	1.881	1.921	1.954	1.982	2.007	2.028	2.047	2.064	2.357	.18
	.305	*.397*	*.468*	*.523*	*.568*	*.606*	*.637*	*.664*	*.687*	*.707*	*.725*	*.741*	*.754*	*.767*	*1.000*	
.19	1.296	1.474	1.596	1.685	1.754	1.809	1.853	1.890	1.921	1.948	1.971	1.991	2.009	2.024	2.294	.19
	.319	*.413*	*.484*	*.540*	*.585*	*.621*	*.652*	*.679*	*.701*	*.721*	*.738*	*.753*	*.767*	*.779*	*1.000*	
.20	1.291	1.464	1.581	1.667	1.732	1.784	1.826	1.861	1.890	1.915	1.936	1.955	1.972	1.987	2.236	.20
	.333	*.429*	*.500*	*.556*	*.600*	*.636*	*.667*	*.692*	*.714*	*.733*	*.750*	*.765*	*.778*	*.789*	*1.000*	
.21	1.286	1.454	1.567	1.648	1.711	1.760	1.800	1.833	1.860	1.884	1.904	1.922	1.937	1.951	2.182	.21
	.347	*.444*	*.515*	*.571*	*.615*	*.650*	*.680*	*.705*	*.727*	*.745*	*.761*	*.776*	*.788*	*.799*	*1.000*	
.22	1.280	1.443	1.552	1.631	1.690	1.737	1.775	1.806	1.832	1.854	1.873	1.890	1.904	1.917	2.132	.22
	.361	*.458*	*.530*	*.585*	*.629*	*.664*	*.693*	*.717*	*.738*	*.756*	*.772*	*.786*	*.798*	*.809*	*1.000*	
.23	1.275	1.433	1.538	1.614	1.671	1.715	1.751	1.780	1.805	1.826	1.844	1.859	1.873	1.885	2.085	.23
	.374	*.473*	*.544*	*.599*	*.642*	*.676*	*.705*	*.729*	*.749*	*.767*	*.782*	*.795*	*.807*	*.818*	*1.000*	
.24	1.270	1.424	1.525	1.597	1.651	1.694	1.728	1.756	1.779	1.799	1.816	1.830	1.843	1.855	2.041	.24
	.387	*.486*	*.558*	*.612*	*.655*	*.689*	*.716*	*.740*	*.759*	*.776*	*.791*	*.804*	*.816*	*.826*	*1.000*	
.25	1.265	1.414	1.512	1.581	1.633	1.673	1.706	1.732	1.754	1.773	1.789	1.803	1.815	1.826	2.000	.25
	.400	*.500*	*.571*	*.625*	*.667*	*.700*	*.727*	*.750*	*.769*	*.786*	*.800*	*.812*	*.824*	*.833*	*1.000*	

r/ab	2	3	4	5	6	7	8	9	10	11	12	13	14	15	∞	r/ab
.26	1.260	1.405	1.499	1.566	1.615											.26

r_AB	2	3	4	5	6	7	8	9	10	11	12	13	14	15	∞	r_AB
.25	1.265 .400	1.414 .500	1.512 .571	1.581 .625	1.633 .667	1.673 .700	1.706 .727	1.732 .750	1.754 .769	1.773 .786	1.789 .800	1.803 .812	1.815 .824	1.826 .833	2.000 1.000	.25
.26	1.260 .413	1.405 .513	1.499 .584	1.566 .637	1.615 .678	1.654 .711	1.684 .738	1.709 .760	1.730 .778	1.748 .794	1.763 .808	1.776 .820	1.788 .831	1.798 .841	1.961 1.000	.26
.27	1.255 .425	1.396 .526	1.487 .597	1.550 .649	1.598 .689	1.635 .721	1.664 .747	1.688 .769	1.707 .787	1.724 .803	1.739 .816	1.751 .828	1.762 .838	1.771 .847	1.925 1.000	.27
.28	1.250 .438	1.387 .538	1.474 .609	1.536 .660	1.581 .700	1.616 .731	1.644 .757	1.667 .778	1.685 .795	1.701 .811	1.715 .824	1.727 .835	1.737 .845	1.746 .854	1.990 1.000	.28
.29	1.245 .450	1.378 .551	1.463 .620	1.521 .671	1.565 .710	1.598 .741	1.625 .766	1.646 .786	1.664 .803	1.679 .818	1.692 .831	1.703 .842	1.713 .851	1.722 .860	1.857 1.000	.29
.30	1.240 .462	1.369 .562	1.451 .632	1.508 .682	1.549 .720	1.581 .750	1.606 .774	1.627 .794	1.644 .811	1.658 .825	1.671 .837	1.681 .848	1.690 .857	1.698 .865	1.826 1.000	.30
.31	1.236 .473	1.361 .574	1.440 .642	1.494 .692	1.534 .729	1.564 .759	1.589 .782	1.608 .802	1.624 .818	1.638 .832	1.650 .844	1.660 .854	1.668 .863	1.676 .871	1.796 1.000	.31
.32	1.231 .485	1.353 .585	1.429 .653	1.481 .702	1.519 .738	1.548 .767	1.571 .790	1.590 .809	1.605 .826	1.618 .838	1.629 .850	1.639 .860	1.647 .868	1.654 .876	1.768 1.000	.32
.33	1.226 .496	1.344 .596	1.418 .663	1.468 .711	1.505 .747	1.533 .775	1.555 .798	1.572 .816	1.587 .831	1.599 .844	1.610 .855	1.619 .865	1.627 .873	1.634 .881	1.741 1.000	.33
.34	1.222 .507	1.336 .607	1.407 .673	1.456 .720	1.491 .756	1.517 .783	1.538 .805	1.555 .823	1.569 .837	1.581 .850	1.591 .861	1.600 .870	1.607 .878	1.614 .885	1.715 1.000	.34
.35	1.217 .519	1.328 .618	1.397 .683	1.443 .729	1.477 .764	1.503 .790	1.523 .812	1.539 .829	1.552 .843	1.563 .856	1.573 .866	1.581 .875	1.588 .883	1.594 .890	1.690 1.000	.35
.36	1.213 .529	1.321 .628	1.387 .692	1.431 .738	1.464 .771	1.488 .797	1.508 .818	1.523 .835	1.536 .849	1.546 .861	1.555 .871	1.563 .880	1.570 .887	1.576 .894	1.667 1.000	.36
.37	1.208 .540	1.313 .638	1.377 .701	1.420 .746	1.451 .779	1.474 .804	1.493 .825	1.508 .841	1.520 .855	1.530 .866	1.538 .876	1.546 .884	1.552 .892	1.558 .898	1.644 1.000	.37
.38	1.204 .551	1.306 .648	1.367 .710	1.409 .754	1.438 .786	1.461 .811	1.478 .831	1.493 .847	1.504 .860	1.514 .871	1.522 .880	1.529 .888	1.535 .896	1.541 .902	1.622 1.000	.38
.39	1.200 .561	1.298 .657	1.358 .719	1.398 .762	1.426 .793	1.448 .817	1.465 .836	1.478 .852	1.489 .865	1.498 .876	1.506 .886	1.513 .893	1.519 .900	1.534 .906	1.601 1.000	.39
.40	1.195 .571	1.291 .667	1.348 .727	1.387 .769	1.414 .800	1.435 .824	1.451 .842	1.464 .857	1.474 .870	1.483 .880	1.491 .889	1.497 .897	1.503 .903	1.508 .909	1.581 1.000	.40
.41	1.191 .582	1.284 .676	1.339 .735	1.376 .777	1.403 .807	1.422 .829	1.438 .848	1.450 .862	1.460 .874	1.469 .884	1.476 .893	1.482 .900	1.487 .907	1.492 .912	1.562 1.000	.41
.42	1.187 .592	1.277 .685	1.330 .743	1.366 .784	1.391 .813	1.410 .835	1.425 .853	1.437 .867	1.446 .879	1.454 .888	1.461 .897	1.467 .904	1.472 .910	1.477 .916	1.543 1.000	.42
.43	1.183 .601	1.270 .694	1.322 .751	1.356 .790	1.380 .819	1.398 .841	1.412 .858	1.424 .872	1.433 .883	1.441 .892	1.447 .901	1.453 .907	1.458 .914	1.462 .919	1.525 1.000	.43
.44	1.179 .611	1.263 .702	1.313 .759	1.346 .797	1.369 .825	1.387 .846	1.400 .863	1.412 .876	1.420 .887	1.427 .896	1.433 .904	1.439 .911	1.443 .917	1.447 .922	1.508 1.000	.44
.45	1.174 .621	1.257 .711	1.305 .766	1.336 .804	1.359 .831	1.375 .851	1.388 .867	1.399 .880	1.407 .891	1.414 .900	1.420 .908	1.425 .914	1.430 .920	1.433 .925	1.491 1.000	.45
.46	1.170 .630	1.250 .719	1.296 .773	1.327 .810	1.348 .836	1.364 .856	1.377 .872	1.387 .885	1.395 .895	1.402 .904	1.407 .911	1.412 .917	1.416 .923	1.420 .927	1.474 1.000	.46
.47	1.166 .639	1.244 .727	1.288 .780	1.318 .816	1.338 .842	1.354 .861	1.366 .876	1.375 .889	1.383 .899	1.389 .907	1.395 .914	1.399 .920	1.403 .925	1.407 .930	1.459 1.000	.47
.48	1.162 .649	1.237 .735	1.280 .787	1.309 .822	1.328 .847	1.343 .866	1.355 .881	1.364 .893	1.371 .902	1.377 .910	1.382 .917	1.387 .923	1.391 .928	1.394 .933	1.443 1.000	.48
.49	1.159 .658	1.231 .742	1.273 .794	1.300 .828	1.319 .852	1.333 .871	1.344 .885	1.353 .896	1.360 .906	1.365 .914	1.370 .920	1.375 .926	1.378 .931	1.381 .935	1.429 1.000	.49
.50	1.155 .667	1.225 .750	1.265 .800	1.291 .833	1.309 .857	1.323 .876	1.333 .889	1.342 .900	1.348 .909	1.354 .917	1.359 .923	1.363 .929	1.366 .933	1.369 .938	1.414 1.000	.50
r_AB	2	3	4	5	6	7	8	9	10	11	12	13	14	15	∞	r_AB

FAB	2	3	4	5	6	7	8	9	10	11	12	13	14	15	∞	FAB
.51	1.151	1.219	1.257	1.282	1.300	1.313	1.323	1.331	1.337	1.343	1.347	1.351	1.355	1.357	1.400	.51
	.675	*.757*	*.806*	*.839*	*.862*	*.879*	*.893*	*.904*	*.912*	*.920*	*.926*	*.931*	*.936*	*.940*	*1.000*	
.52	1.147	1.213	1.250	1.274	1.291	1.303	1.313	1.321	1.327	1.332	1.336	1.340	1.343	1.346	1.387	.52
	.684	*.765*	*.812*	*.844*	*.867*	*.883*	*.897*	*.907*	*.915*	*.923*	*.929*	*.934*	*.938*	*.942*	*1.000*	
.53	1.143	1.207	1.243	1.266	1.282	1.294	1.303	1.311	1.316	1.321	1.326	1.329	1.332	1.335	1.374	.53
	.693	*.772*	*.819*	*.849*	*.871*	*.888*	*.900*	*.910*	*.919*	*.926*	*.931*	*.936*	*.940*	*.944*	*1.000*	
.54	1.140	1.201	1.236	1.258	1.273	1.285	1.294	1.301	1.306	1.311	1.315	1.318	1.321	1.324	1.361	.54
	.701	*.779*	*.824*	*.854*	*.876*	*.893*	*.904*	*.914*	*.922*	*.928*	*.934*	*.939*	*.943*	*.946*	*1.000*	
.55	1.136	1.195	1.229	1.250	1.265	1.276	1.284	1.291	1.296	1.301	1 305	1.308	1.311	1.313	1.348	.55
	.710	*.786*	*.830*	*.859*	*.880*	*.895*	*.907*	*.917*	*.924*	*.931*	*.936*	*.941*	*.945*	*.948*	*1.000*	
.56	1.132	1.190	1.222	1.242	1.257	1.267	1.275	1.282	1.287	1.291	1.295	1.298	1.300	1.303	1.336	.56
	.718	*.792*	*.836*	*.864*	*.884*	*.899*	*.911*	*.920*	*.927*	*.933*	*.939*	*.943*	*.947*	*.950*	*1.000*	
.57	1.129	1.184	1.215	1.235	1.248	1.258	1.266	1.272	1.277	1.281	1.285	1.288	1.290	1.292	1.325	.57
	.726	*.799*	*.841*	*.869*	*.888*	*.903*	*.914*	*.923*	*.930*	*.936*	*.941*	*.945*	*.949*	*.952*	*1.000*	
.58	1.125	1.179	1.208	1.227	1.240	1.250	1.257	1.263	1.268	1.272	1.275	1.278	1.280	1.282	1.313	.58
	.734	*.806*	*.847*	*.873*	*.892*	*.906*	*.917*	*.926*	*.932*	*.938*	*.943*	*.947*	*.951*	*.954*	*1.000*	
.59	1.122	1.173	1.202	1.220	1.232	1.242	1.249	1.254	1.259	1.263	1.266	1.268	1.271	1.273	1.302	.59
	.742	*.812*	*.852*	*.878*	*.896*	*.910*	*.920*	*.928*	*.935*	*.941*	*.945*	*.949*	*.953*	*.956*	*1.000*	
.60	1.118	1.168	1.195	1.213	1.225	1.234	1.240	1.246	1.250	1.254	1.257	1.259	1.261	1.263	1.291	.60
	.750	*.818*	*.857*	*.888*	*.900*	*.918*	*.923*	*.931*	*.938*	*.943*	*.947*	*.951*	*.955*	*.957*	*1.000*	
.61	1.115	1.162	1.189	1.206	1.217	1.226	1.232	1.237	1.241	1.245	1.248	1.250	1.252	1.254	1.280	.61
	.758	*.824*	*.862*	*.887*	*.904*	*.916*	*.926*	*.934*	*.940*	*.945*	*.949*	*.953*	*.956*	*.959*	*1.000*	
.62	1.111	1.157	1.183	1.199	1.210	1.218	1.224	1.229	1.233	1.236	1.239	1.241	1.243	1.245	1.270	.62
	.765	*.830*	*.867*	*.891*	*.907*	*.919*	*.929*	*.936*	*.942*	*.947*	*.951*	*.955*	*.958*	*.961*	*1.000*	
.63	1.108	1.152	1.176	1.192	1.202	1.210	1.216	1.221	1.224	1.228	1.230	1.232	1.234	1.236	1.260	.63
	.773	*.836*	*.872*	*.895*	*.911*	*.923*	*.932*	*.939*	*.945*	*.949*	*.953*	*.957*	*.960*	*.962*	*1.000*	
.64	1.104	1.147	1.170	1.185	1.195	1.203	1.208	1.213	1.216	1.219	1.222	1.224	1.226	1.227	1.250	.64
	.780	*.842*	*.877*	*.899*	*.914*	*.926*	*.934*	*.941*	*.947*	*.951*	*.955*	*.959*	*.961*	*.964*	*1.000*	
.65	1.101	1.142	1.164	1.179	1.188	1.195	1.201	1.205	1.208	1.211	1.213	1.215	1.217	1.219	1.240	.65
	.788	*.848*	*.881*	*.903*	*.918*	*.929*	*.937*	*.944*	*.949*	*.953*	*.957*	*.960*	*.963*	*.965*	*1.000*	
.66	1.098	1.137	1.159	1.172	1.181	1.188	1.193	1.197	1.200	1.203	1.205	1.207	1.209	1.210	1.231	.66
	.795	*.853*	*.886*	*.907*	*.921*	*.931*	*.940*	*.946*	*.951*	*.955*	*.959*	*.962*	*.965*	*.967*	*1.000*	
.67	1.094	1.132	1.153	1.166	1.174	1.181	1.186	1.190	1.193	1.195	1.197	1.199	1.201	1.202	1.222	.67
	.802	*.859*	*.890*	*.910*	*.924*	*.934*	*.942*	*.948*	*.953*	*.957*	*.961*	*.963*	*.966*	*.968*	*1.000*	
.68	1.091	1.127	1.147	1.159	1.168	1.174	1.179	1.182	1.185	1.188	1.190	1.191	1.193	1.194	1.213	.68
	.810	*.864*	*.895*	*.914*	*.927*	*.937*	*.944*	*.950*	*.955*	*.959*	*.962*	*.965*	*.967*	*.970*	*1.000*	
.69	1.088	1.123	1.141	1.153	1.161	1.167	1.171	1.175	1.178	1.180	1.182	1.184	1.185	1.186	1.204	.69
	.817	*.870*	*.899*	*.918*	*.930*	*.940*	*.947*	*.952*	*.957*	*.961*	*.964*	*.967*	*.969*	*.971*	*1.000*	
.70	1.085	1.118	1.136	1.147	1.155	1.160	1.164	1.168	1.170	1.173	1.174	1.176	1.177	1.179	1.195	.70
	.824	*.875*	*.903*	*.921*	*.933*	*.942*	*.949*	*.955*	*.959*	*.962*	*.966*	*.968*	*.970*	*.972*	*1.000*	
.71	1.081	1.113	1.130	1.141	1.148	1.154	1.158	1.161	1.163	1.165	1.167	1.169	1.170	1.171	1.187	.71
	.830	*.880*	*.907*	*.924*	*.936*	*.945*	*.951*	*.957*	*.961*	*.964*	*.967*	*.970*	*.972*	*.973*	*1.000*	
.72	1.078	1.109	1.125	1.135	1.142	1.147	1.151	1.154	1.156	1.158	1.160	1.161	1.162	1.164	1.179	.72
	.837	*.885*	*.911*	*.928*	*.939*	*.947*	*.954*	*.959*	*.963*	*.966*	*.969*	*.971*	*.973*	*.975*	*1.000*	
.73	1.075	1.104	1.120	1.129	1.136	1.141	1.144	1.147	1.149	1.151	1.153	1.154	1.155	1.156	1.170	.73
	.844	*.890*	*.915*	*.931*	*.942*	*.950*	*.956*	*.961*	*.964*	*.967*	*.970*	*.972*	*.974*	*.976*	*1.000*	
.74	1.072	1.100	1.115	1.124	1.130	1.134	1.138	1.140	1.143	1.144	1.146	1.147	1.148	1.149	1.162	.74
	.851	*.895*	*.919*	*.934*	*.945*	*.953*	*.958*	*.962*	*.966*	*.969*	*.972*	*.974*	*.976*	*.977*	*1.000*	
.75	1.069	1.095	1.109	1.118	1.124	1.128	1.131	1.134	1.136	1.138	1.139	1.140	1.141	1.142	1.155	.75
	.857	*.900*	*.923*	*.938*	*.947*	*.955*	*.960*	*.964*	*.968*	*.971*	*.973*	*.975*	*.977*	*.978*	*1.000*	

FAB	2	3	4	5	6	7	8	9	10	11	12	13	14	15	∞	FAB
.76	1.066	1.091	1.104	1.112	1.118	1.122	1.125	1.127	1.129	1.131	1.132	1.133	1.134	1.135	1.147	.76
	.864	*.905*	*.927*	*.941*	*.950*	*.957*	*.962*	*.966*	*.969*	*.972*	*.974*	*.976*	*.978*	*.979*	*1.000*	

FAR	2	3	4	5	6	7	8	9	10	11	12	13	14	15	∞	FAR
.75	1.069 *.857*	1.095 *.900*	1.108 *.923*	1.118 *.938*	1.125 *.947*	1.131 *.955*	1.135 *.960*	1.139 *.964*	1.142 *.968*	1.144 *.971*						
FAR	2	3	4	5	6	7	8	9	10	11	12	13	14	15	∞	FAR
.76	1.066 *.864*	1.091 *.906*	1.104 *.927*	1.112 *.941*	1.118 *.950*	1.122 *.957*	1.125 *.962*	1.127 *.966*	1.129 *.969*	1.131 *.972*	1.132 *.974*	1.133 *.976*	1.134 *.978*	1.135 *.979*	1.147 *1.000*	.76
.77	1.063 *.870*	1.087 *.909*	1.099 *.931*	1.107 *.944*	1.112 *.953*	1.116 *.959*	1.119 *.964*	1.121 *.968*	1.123 *.971*	1.124 *.974*	1.126 *.976*	1.127 *.978*	1.128 *.979*	1.128 *.980*	1.140 *1.000*	.77
.78	1.060 *.876*	1.083 *.914*	1.094 *.934*	1.102 *.947*	1.107 *.955*	1.110 *.961*	1.113 *.966*	1.115 *.970*	1.117 *.973*	1.118 *.975*	1.119 *.977*	1.120 *.979*	1.121 *.980*	1.122 *.982*	1.132 *1.000*	.78
.79	1.057 *.883*	1.078 *.919*	1.089 *.938*	1.096 *.950*	1.101 *.958*	1.104 *.963*	1.107 *.968*	1.109 *.971*	1.110 *.974*	1.112 *.976*	1.113 *.978*	1.114 *.980*	1.115 *.981*	1.115 *.983*	1.125 *1.000*	.79
.80	1.054 *.889*	1.074 *.923*	1.085 *.941*	1.091 *.952*	1.095 *.960*	1.099 *.966*	1.101 *.970*	1.103 *.973*	1.104 *.976*	1.106 *.978*	1.107 *.980*	1.107 *.981*	1.108 *.982*	1.109 *.984*	1.118 *1.000*	.80
.81	1.051 *.895*	1.070 *.927*	1.080 *.945*	1.086 *.955*	1.090 *.962*	1.093 *.968*	1.095 *.972*	1.097 *.976*	1.098 *.977*	1.099 *.979*	1.100 *.981*	1.101 *.982*	1.102 *.984*	1.103 *.985*	1.111 *1.000*	.81
.82	1.048 *.901*	1.066 *.932*	1.075 *.948*	1.081 *.958*	1.085 *.965*	1.087 *.970*	1.089 *.973*	1.091 *.976*	1.092 *.979*	1.093 *.980*	1.094 *.982*	1.095 *.983*	1.096 *.985*	1.096 *.986*	1.104 *1.000*	.82
.83	1.045 *.907*	1.062 *.936*	1.071 *.951*	1.076 *.961*	1.079 *.967*	1.082 *.972*	1.084 *.975*	1.085 *.978*	1.087 *.980*	1.088 *.982*	1.088 *.983*	1.089 *.984*	1.090 *.986*	1.090 *.987*	1.098 *1.000*	.83
.84	1.043 *.913*	1.058 *.940*	1.066 *.956*	1.071 *.963*	1.074 *.969*	1.077 *.974*	1.078 *.977*	1.080 *.979*	1.081 *.981*	1.082 *.983*	1.083 *.984*	1.083 *.986*	1.084 *.987*	1.084 *.987*	1.091 *1.000*	.84
.85	1.040 *.919*	1.054 *.944*	1.061 *.958*	1.066 *.966*	1.069 *.971*	1.071 *.975*	1.073 *.978*	1.074 *.981*	1.075 *.983*	1.076 *.984*	1.077 *.986*	1.077 *.987*	1.078 *.988*	1.078 *.988*	1.085 *1.000*	.85
.86	1.037 *.925*	1.050 *.949*	1.057 *.961*	1.061 *.968*	1.064 *.974*	1.066 *.977*	1.068 *.980*	1.069 *.982*	1.070 *.984*	1.070 *.985*	1.071 *.987*	1.072 *.988*	1.072 *.989*	1.073 *.989*	1.078 *1.000*	.86
.87	1.034 *.930*	1.046 *.953*	1.053 *.964*	1.056 *.971*	1.059 *.976*	1.061 *.979*	1.062 *.982*	1.063 *.984*	1.064 *.985*	1.065 *.987*	1.065 *.988*	1.066 *.989*	1.066 *.989*	1.067 *.990*	1.072 *1.000*	.87
.88	1.031 *.936*	1.043 *.957*	1.048 *.967*	1.052 *.973*	1.054 *.978*	1.056 *.981*	1.057 *.983*	1.058 *.985*	1.059 *.987*	1.059 *.988*	1.060 *.989*	1.060 *.990*	1.061 *.990*	1.061 *.991*	1.066 *1.000*	.88
.89	1.029 *.942*	1.039 *.960*	1.044 *.970*	1.047 *.976*	1.049 *.980*	1.051 *.983*	1.052 *.985*	1.053 *.986*	1.054 *.988*	1.054 *.989*	1.055 *.990*	1.055 *.991*	1.055 *.991*	1.056 *.992*	1.060 *1.000*	.89
.90	1.026 *.947*	1.035 *.964*	1.040 *.973*	1.043 *.978*	1.044 *.982*	1.046 *.984*	1.047 *.986*	1.048 *.988*	1.048 *.989*	1.049 *.990*	1.049 *.991*	1.050 *.992*	1.050 *.992*	1.050 *.993*	1.054 *1.000*	.90
.91	1.023 *.953*	1.031 *.968*	1.036 *.976*	1.038 *.981*	1.040 *.984*	1.041 *.986*	1.042 *.988*	1.043 *.989*	1.043 *.990*	1.044 *.991*	1.044 *.992*	1.044 *.992*	1.045 *.993*	1.045 *.993*	1.048 *1.000*	.91
.92	1.021 *.958*	1.028 *.972*	1.031 *.979*	1.034 *.983*	1.035 *.986*	1.036 *.988*	1.037 *.989*	1.038 *.990*	1.038 *.991*	1.038 *.992*	1.039 *.993*	1.039 *.993*	1.039 *.994*	1.040 *.994*	1.043 *1.000*	.92
.93	1.018 *.964*	1.024 *.976*	1.027 *.982*	1.029 *.985*	1.031 *.988*	1.031 *.989*	1.032 *.991*	1.033 *.992*	1.033 *.993*	1.033 *.993*	1.034 *.994*	1.034 *.994*	1.034 *.995*	1.034 *.995*	1.037 *1.000*	.93
.94	1.015 *.969*	1.021 *.979*	1.023 *.984*	1.025 *.987*	1.026 *.989*	1.027 *.991*	1.027 *.992*	1.028 *.993*	1.028 *.994*	1.028 *.994*	1.029 *.995*	1.029 *.995*	1.029 *.995*	1.029 *.996*	1.031 *1.000*	.94
.95	1.013 *.974*	1.017 *.983*	1.019 *.987*	1.021 *.990*	1.022 *.991*	1.022 *.993*	1.023 *.993*	1.023 *.994*	1.023 *.995*	1.024 *.995*	1.024 *.996*	1.024 *.996*	1.024 *.996*	1.024 *.997*	1.026 *1.000*	.95
.96	1.010 *.980*	1.014 *.986*	1.015 *.990*	1.016 *.992*	1.017 *.993*	1.018 *.994*	1.018 *.995*	1.018 *.995*	1.019 *.996*	1.019 *.996*	1.019 *.997*	1.019 *.997*	1.019 *.997*	1.019 *.997*	1.021 *1.000*	.96
.97	1.008 *.985*	1.010 *.990*	1.011 *.992*	1.012 *.994*	1.013 *.995*	1.013 *.996*	1.013 *.996*	1.014 *.997*	1.014 *.997*	1.014 *.997*	1.014 *.997*	1.014 *.998*	1.014 *.998*	1.014 *.998*	1.015 *1.000*	.97
.98	1.005 *.990*	1.007 *.993*	1.008 *.995*	1.008 *.996*	1.008 *.997*	1.009 *.997*	1.009 *.997*	1.009 *.998*	1.009 *.998*	1.009 *.998*	1.009 *.998*	1.009 *.998*	1.009 *.999*	1.009 *.999*	1.010 *1.000*	.98
.99	1.003 *.995*	1.003 *.997*	1.004 *.997*	1.004 *.998*	1.004 *.998*	1.004 *.999*	1.004 *.999*	1.004 *.999*	1.005 *.999*	1.005 *.999*	1.005 *.999*	1.005 *.999*	1.005 *.999*	1.005 *.999*	1.005 *1.000*	.99
1.00	1.000 *1.000*	1.000 *1.000*	1.000 *1.000*	1.000 *1.000*	1.000 *1.000*	1.000 *1.000*	1.000 *1.000*	1.000 *1.000*	1.000 *1.000*	1.000 *1.000*	1.000 *1.000*	1.000 *1.000*	1.000 *1.000*	1.000 *1.000*	1.000 *1.000*	1.00
FAR	2	3	4	5	6	7	8	9	10	11	12	13	14	15	∞	FAR

other to the extent of .65, the examinations are too short,—contain too few questions or scoring points or have too short a time limit; the table shows that a minimally acceptable reliability of, say, .85 would require in this case the examinations to contain three times as many questions, or items. In test construction, for example, tests of low initial reliability and fair validity may often be salvaged by the simple expedient of making them longer. The table makes this decision a very easy matter. Many of the published tables of intercorrelation coefficients are valueless for the reason that the tests used were so short that a repetition of the experiment might yield almost any other values than those published. The authors propose, for purposes of discussion, that no intercorrelation tables should ordinarily be presented unless based on tests or units (those on which the intercorrelations are based), which have at least a reliability coefficient of .90, unless the reliability coefficients are given, or unless the nature of the experiment precludes the possibility of such reliability being attained. Such a standard, if used in the past would have saved, for example, our being assailed in the literature by statements in regard to "the correlation between mechanical ability and intelligence" ranging from, say,—.60 to +.60. Such a consideration, again, for example, would require, in general, that we repeat the majority of psychomotor tests many times and average the individual's several scores in order to obtain a reliable score for the individual.

A COMPARISON OF TWO METHODS OF LEARNING

R. D. RUSSELL
University of Idaho

IT has been said that there are four methods of approach to learning:

1. Studying the thing itself;
2. Studying a picture of the thing;
3. Hearing about it;
4. Reading about it.

Within the last few years Freeman, McCluskey, James, Weber and others have made extended investigations regarding the value of pictures in teaching. Only incidentally have they considered the relative efficiency of learning by lecture and reading the material. The writer is not informed of any comparisons having been made between learning by reading and by hearing material read.

A number of investigators have attempted to determine whether the visual or auditory method of presentation gives a better result in the recall of the material presented. Hawkins[1] presented words visually and orally. The object of the experiment was to determine which was the stronger, auditory or visual memory. His results show that in the primary and grammar schools students recall more from auditorially presented material, while in the high school and college the visual method has an advantage.

Whitehead[2] presented nonsense syllables visually and orally to adults. He found the visual method superior in memorizing but material memorized orally was retained slightly better. Calkins[3] experimented with college students. Words were presented visually and orally. She found a superiority in favor of the visual method.

Pohlmann[4] studied the effect of visual, auditory, visual-auditory and visual-auditory-motor presentations of objects, words, non-sense syllables and numbers, on school children from 9 to 14 years of age. He found that auditory presentation is better than visual with significant material, but that visual presentation is better with nonsense material.

The studies referred to above make use of nonsense material, numbers, and words. The present study differs from them in that it makes use of context material. Several studies have been made, however, of learning by lecture and by reading.

[1] Hawkins, Chauncey J. "Experiments on Memory Types." *Psychological Review.* (Vol. 4, 1897), pp. 289-294.

[2] Whitehead, Louis Grant. "A Study of Visual and Aural Memory Processes." *Psychological Review.* (Vol. 3, 1896), pp. 259-269.

[3] Calkins, Mary Whiton. "Short Studies in Memory and in Association from Wellesley College." *Psychological Review.* (Vol. 5, 1898), pp. 451-462.

[4] Pohlmann, A. *Experimentelle Beitrage zur Lehre vom Gedachtnis.* 1905.

Lacy[5] made a study in which he found the oral presentation superior to the visual. His subjects were 7th, 8th and 9th grade children.

Weber[6] has recently made a study of the relative effectiveness of four methods of presentation, namely, study, teacher, film and film-lecture. He compared the relative effect of the lecture by the teacher to the students with study by the pupils themselves. The students who read the lesson for themselves made on the average of .6 of a point better than the students to whom the teacher lectured. The experiment was with 7th grade students.

James[7] has recently made a very comprehensive study involving six different methods of presentation. He found in the 8th grade a superiority in favor of reading.

This study differs from those of Lacy, Weber and James in that the material was read to the students instead of given in the form of a lecture. It is hoped that this study will give some suggestion as to the most economical period for the introduction of book instruction in the elementary school. Johnson[8] states regarding history that:

> "Below the 7th grade the ability of the average pupil to comprehend is so far in advance of his ability to read that the necessary details are most economically and efficiently supplied by the teacher. Beginning with the 7th grade required readings outside of the text may, to an ever increasing extent, be substituted by contributions of the teacher."

The material used in this investigation was a composition of about 1000 words on the Mongoose. The material was used concerning this animal because it, in all probability, would not duplicate the work of the children in school, and further because it represented a type of material common to schoolroom use.

METHOD OF INVESTIGATION

The National Intelligence Test was given to 1080 pupils in the 5th, 7th and 9th grades. The students in each grade were then divided into three groups of equal ability as measured by this test. Due to a loss in grouping this number was reduced to 690.

The material was presented by the regular class-room teacher in three different ways. In one case it was read to the students twice in

[5] Lacy, John V. "The Relative Value of Motion Pictures as an Educational Agency." *Teachers College Record.* (Vol. 20, November, 1919), pp. 453-465.

[6] Weber, Joseph J. *Comparative Effectiveness of Some Visual Aids in Seventh Grade Instruction.* (The Educational Screen, Inc., 1922), p. 90.

[7] James, Hadden W. "Relative Effectiveness of Six Forms of Lesson presentation: Film, Lecture, Still-Picture, Film-Picture, Film-Music, and Reading with Particular Emphasis on the Efficiency of Different Types of Material for Film Presentation." An unpublished thesis in the Department of Education, State University of Iowa, 1923, page 241.

[8] Johnson, Henry. *Teaching of History.* 1915, page 287.

succession. The second procedure was to give the students the material and ask them to read it at their normal rate, giving them the same amount of time for reading as the teacher took in the first method in reading it to them. The third group read the material keeping the number of impressions constant with that of the first group without regard to time. This group was simply told to read the material through twice at their normal rate and as soon as they finished to stop and raise their hands.

After the presentation of the material the pupils were given two tests, one of the essay type and the other the true-false, the essay type being administered first. The scores on these tests were combined on the basis of the ratio of their true standard deviations to facilitate comparison of groups. After pairing the scores in Grade 5 only 72 pupils remained in each of the three groups. Following are the averages of the combined scores:

GRADE 5

		Teachers Reading	Pupils Reading Time Constant	Pupils Reading Impressions Constant
		(1)	(2)	(3)
Means		57.29	49.58	53.95
S.D.'s		16.329	14.754	17.079
S.D.M.'s		1.926	1.74	2.016
S.D.	m_1-m_2	2.595		
S.D.	m_1-m_3	2.788		

In comparing the scores made when the teacher read to the pupils and when the pupils read the material for themselves in the length of time required of the teacher to read, it will be observed that the mean of column 1 is 7.71 points higher than that of column 2. This difference (7.71) is 2.97 times the standard deviation of the difference which means that the chances are 332 to 1 that the difference represents a real superiority of group one over group two and is not to be accounted for by the mere operation of chance.

A comparison of columns one and three shows also a distinct superiority in favor of the method of reading the material to the pupils.

After grouping 104 students remained in each of the three groups of Grade 7. The means for these groups are as follows:

GRADE 7

		Teachers Reading	Pupils Reading Time Constant	Pupils Reading Impressions Constant
		(1)	(2)	(3)
Means		75.79	76.04	75.43
S.D.'s		17.187	16.539	17.844
S.D. M. 's		1.685	1.622	1.749
S.D.	m_1-m_2	2.328		
S.D.	m_1-m_3	2.428		

It will be observed that the differences in the means of columns 1, 2 and 3 are negligible in Grade 7. They are no greater than chance would allow. However, the students represented by column 2 read the material 2.4 times. Although the time was constant and the results practically the same it may be possible that the group having the greater number of impressions has a slight advantage.

In comparing columns 1 and 3 it will be noticed that no significant difference is apparent. The students represented in column 3 read the material over twice in a slightly shorter period of time than the teacher took in reading to the students represented in column 1.

Three groups of 51 students each were represented in Grade 9. The results follow:

GRADE 9

		TEACHERS READING	PUPILS READING TIME CONSTANT	PUPILS READING IMPRESSIONS CONSTANT
		(1)	(2)	(3)
Means		85.41	86.77	87.88
S.D.'s		12.576	12.999	15.087
S.D.M.'s		1.761	1.821	2.113
S.D.	m_1-m_2	2.533		
S.D.	m_1-m_3	2.751		

A comparison of columns 1 and 2 of Grade 9 shows a difference of 1.36 points in favor of the students reading the material themselves. This difference, however, is not significant in as much as the Standard Deviation of the difference is 2.5. A record of the number of times the students read the material in column 1 was not kept.

Column 3, in which the students read the material twice to compare with the number of impressions of those in column 1, shows a superiority over column 1.

The result of the experiment in Grade 9 indicates possibly that students learn more from reading themselves than from hearing someone else read when the time and the number of impressions are held constant.

CONCLUSIONS

The result of this investigation seems to justify the conclusions that:

1. Students learn more from having material read to them than in reading it themselves in Grade 5.

2. The relative effectiveness of the two methods is practically equal in Grade 7.

3. The advantage is slightly in favor of the method of learning by reading in Grade 9.

AN EXPERIMENTAL INVESTIGATION OF THE RELATIVE EFFECTIVENESS OF TWO PLANS OF SUPERVISED STUDY

HARL R. DOUGLASS
University of Pennsylvania

THE chief purpose of the group of experiments described in the following paragraphs was to determine the relative efficiency of the S-R sequence and the R-S sequence in the administration of supervised study. By "S-R sequence" is meant nothing more than having supervised study precede the recitation in the class period and by "R-S sequence" that the recitation precedes the supervised study.

Many progressive teachers employing supervised study do not attempt to divide each class period into two portions—supervised study and recitation—but prefer to plan in larger than daily units covering a number of days or even weeks, e. g., challenges, projects, contracts, etc., and the majority of those nationally eminent in the field of philosophy and practice of teaching lean heavily in the direction of these larger units. Yet it is quite likely that in the majority of classrooms where some form of supervised study is used regularly and in which the lengthened period is available, daily or almost daily recitations are in vogue. The fact that this study is based on the latter practice should not be taken as an indication that the author does not favor some form of larger units for those teachers who are able to plan them.

Past practice has very largely favored the R-S sequence—study following recitation. It has been advanced that the recitation should begin where the discussion left off and should follow assignment as closely as possible. It has been claimed that the study of material to be recited immediately after is not as likely to result in permanent learning as study with a view to retention for at least 23½ hours. In connection with this claim, it is offered, particularly by those who think of teaching as lesson learning and testing, that a recitation immediately following study is not as adequate a test of learning as one held on the following day as would be the case where the R-S sequence is employed.

Among other things the numerous attempts to realize in practice the theoretical advantages of supervised study have brought to light the two following factors which have operated to seriously impair its efficiency: (1) in the supervised study time following a recitation, pupils are not inclined to take the responsibility for study as seriously as necessary for effectiveness; they prefer to study for recitations looming ahead in the

next or subsequent periods or to postpone study until a time when the recitation is more imminent; (2) where the R-S sequence is used more often than not the recitation is permitted to run over into time theoretically allotted to study. Largely because of these two developments the lengthened period has not proven in general to be much if any improvement over the short period where no regular supervised study is contemplated despite the increased cost of supervised study due to the consequent lightened teacher's load.

There have been those who have claimed that employment of the S-R sequence that the contribution to increased effectiveness resulting from placing supervised study just before recitation thereby utilizing the increased interests of pupils and by making certain that the time theoretically allotted to supervised study is actually realized, will more than offset the disadvantages accruing from employing the S-R sequence. It was the purpose of this group of experiments to furnish some objective experimental data pertinent to this question.

THE EXPERIMENTAL GROUPS AND TECHNIQUE

The experimental subjects were students in the University of Oregon High School, a six-year junior-senior high school. They pay no fees and are admitted in the order of application. The regular faculty is composed of somewhat superior, progressive, experienced teachers comparable to the professionally alive teachers in any city high school where a reasonably adequate salary schedule obtains.

Ten pairs of classes were employed as follows:

History J2—(second semester, seventh grade, American history)
Civics J4—(second semester, eighth grade, community civics)
History J6—(second semester, ninth grade, medieval history)
History S2—(second semester, tenth grade, modern European history)
Civics S4—(second semester, eleventh grade, U. S. government)
English J6—(second semester, ninth grade, literature)
Mathematics J2—(second semester, seventh grade, unified mathematics)
Mathematics J4—(second semester, eighth grade, unified mathematics)
Mathematics J6—(second semester, ninth grade, unified mathematics)
Science J4—(second semester, eighth grade, general science)

The students were paired on the following points: Chronological age, Otis Intelligence Test scores (advanced examination), and an achievement test intended to furnish a measure of special aptitude.

The subject matter tests employed were as follows: All history and civics classes, the *Gregory American History Test*, Part III, for seventh and eighth grades and high school; English J6, the *Monroe Silent Read-*

ing Test; Mathematics J2, *Stanford Achievement Arithmetic Examinations;* Mathematics J4 and J6, *the Douglass Standard Diagnostic Algebra Tests;* Science J4, the *Ruch-Popenoe General Science Test.*

Owing to the great difficulty in arranging adequate pairs on three different criteria, composite scores were calculated from the Otis scores and the achievement test scores in which the weightings were equalized by the method of equal variabilities, i. e., the scores on one of the two tests were multiplied by the ratio of the standard deviation of the scores made on the other test by the pupils of the class to be paired to the standard deviation of the scores made on the test the scores of which were being converted.

The two experimental sections were so constructed from these pairings as to have one of each pair in each section and to result in sections of equal mean chronological age, composite score, and initial test score[1] and of approximately equal variability in chronological age and composite score.

CONTROLLING ALL BUT THE EXPERIMENTAL VARIABLE

Each pair of sections recited at the same hour and in equally favorable rooms. The teachers assigned to these sections were selected so as to be of approximately equal prospective ability and were exchanged at the middle of the experimental period of ten weeks the influence of teacher efficiency being thus equalized between the two sections. Careful logs of the experiments, including absences and interruptions were kept. Because of the fact that each pair of sections recited at the same time, the only interfering circumstances affecting any pair of sections unequally were absences and withdrawals. In these cases of withdrawals and excessive absences of more than five days, the absent pupil was dropped from experimental consideration and the pupil paired with him either paired with some pupil in his section not originally paired or dropped also from consideration.

MEASUREMENT OF RESULTS

Objective tests requiring from 80 to 110 minutes were prepared to cover the work to be taught during the experimental periods. These initial tests were given the two days just preceding the beginning of the experiment. Similar, but not strictly comparable, objective tests were devised and given as final tests. These tests included such types of items

[1] To be discussed later.

as completion, true-false, multiple-choice, matching, analogy, and enumeration exercises. In most instances the final tests included at least a few items taken from the initial tests. The initial and final tests for literature were identical. In the mathematics J6 sections, two forms of the Douglass Algebra tests were employed as initial and final tests. The reliabilities computed from scores of the half-tests of the students taking them by the Spearman-Brown prophecy formula ranged from .61±.09 to .94±.01 the middle 60 percent of them ranging from .75±.06 to .90 ±.02.

As a measure of the relative gain of the two sections was taken the following difference Σ(FT — IT) of one section — Σ(FT — IT) of the other. This of course is equal to the difference of the sums of the final test scores of the two sections since the sections were so constructed that the sums (or means) of the initial test scores of the two sections were equal.

THE EXPERIMENTAL RESULTS

The most important experimental results are summarized in the table given below.

TABLE I

CHANCES IN 100 THAT SUPERIORITY IS IN THE DIRECTION OF THE OBTAINED SUPERIORITY

	d	SD_d	$\frac{d}{SD}$		Chances in 100
History J2	.56	1.89	.30	SR > RS	62
History J4	.70	1.97	.36	SR > RS	64
History J6	3.39	3.92	.86	SR > RS	81
History S2	4.48	3.22	1.39	SR > RS	92
History S4	3.92	1.95	2.01	SR > RS	95
English J6	5.62	2.27	2.48	SR > RS	99
Mathematics J2	1.50	.86	1.83	RS < SR	97
Mathematics J4	.24	1.56	.16	RS < SR	56
Mathematics J6	1.14	.96	1.19	RS < SR	88
Science J4	4.38	5.11	.86	RS < SR	81

d = the difference in mean gain of the two sections, i.e., $\Sigma(FT_1 - FT_2) \div N$

SD_d = the standard deviation of the difference in mean gain, as given by the following formula:

$$SD_d = \sqrt{SD^2_{M1} + SD^2_{M2} - 2r_{M1\ M2}\ SD_{M1}\ SD_{M2}}$$

$\frac{d}{SD}$ = the difference expressed in terms of its standard deviation.

In the last two columns of this table shown for each pair of sections was superior and what the chances are (by reference to some normal integral table) that the obtained difference is significant, i. e., that the apparent superiority of the section as indicated was not due to chance or sampling errors.

Due to the fact that the initial and final test units were not comparable, that is, their relative values are not known, no comparisons can be made, such as will enable us to say that one sequence produced results so

many percent better than the other, except in two instances. In the literature sections, where the same test was used as initial and as final tests and in the sections in mathematics J6 section were two comparable forms of the Douglass Algebra tests were used, the following comparisons may be made.

TABLE II

PERCENTAGES OF SUPERIORITY OF THE MORE EFFECTIVE SEQUENCE FOR ENGLISH J6 AND MATHEMATICS J6

	GAIN IN S-R SECTION	GAIN IN R-S SECTION	SUPERIOR GAIN	PERCENTAGE OF SUPERIORITY
English J6	699.6	626.6	70	11.17
Mathematics J6	150.8	166.6	16	9.59

While no exact statement can be made as to the percentages of superiority of the more effective sequence in classes other than those above, in view of the relative chances among the different classes that the obtained superiority of the apparently superior sequence is positive and in view of the fact that they are based upon groups of students of somewhat similar numbers, it is not unreasonable to think of the percentages of superiority as ranging from materially less than 10 percent (in the cases of mathematics J4, and history J4) to not much more than 10 percent. The upper limit seems reasonable in view of the fact that the percentage of superiority in the case of the English scores made on the same test was only 11.17 percent in spite of the fact that the superiority of the more effective sequence was established as not being due to chance more clearly than in the case of any other class.

It was interesting to note the relationship of the opinions of the teachers in charge of the experimental sections with reference to the relative effectiveness of the two sequences to the actual experimental results. Table III below is pertinent to this point.

TABLE III

COMPARISON OF EXPERIMENTAL RESULTS WITH TEACHERS' JUDGMENT BY CLASSES

	EXPERIMENTAL RESULTS FAVORING		TEACHER JUDGMENTS FAVORING		
Class	R-S	S-R	R-S	neither	S-R
History J2		x	1	1	
History J4		x	2		
History J6		x	1		1
History S2		x	2		
History S4		x	1		1
English J6		x	1	1	
Mathematics J2	x		1		
Mathematics J4	x			2	
Mathematics J6	x		1		1
Science J4	x		1	1	

Table IV furnishes evidence as to the relative effect of the two sequences upon variability of progress within the section. This table should be interpreted similarly to Table I.

TABLE IV

VARIATION OF GAINS OF PUPILS BY CLASS AND BY SEQUENCE

Class	SD_{gain} S-R	SD_{gain} R-S	SD_1-SD_2	SD_{diff}	$\frac{SD}{SD_{diff}}$	Chances that difference is real
History J2	7.36	7.76	—.46	1.67	.24	60 to 40
History J4	5.70	5.09	.61	1.35	.46	68 to 32
History J6	10.00	10.04	—.04	2.52	.02	51 to 49
History S2	8.05	8.01	.04	2.32	.02	51 to 49
History S4	6.27	5.83	.44	1.59	.28	61 to 39
English J6	5.11	6.24	—1.13	1.57	.72	76 to 24
Math. J2	3.72	3.84	—.12	.73	.16	56 to 44
Math. J4	3.32	4.01	—1.29	1.17	1.10	86 to 14
Math. J6	4.34	3.08	1.26	1.16	1.09	86 to 14
Science J4	16.38	13.86	2.52	3.88	.68	75 to 25

The relative effectiveness of the sequences for students of varying intelligence was investigated. Initial and final test scores of all pupils were expressed in terms of their own standard deviations. The mean relative superiority of the R-S Sections were each computed in these terms. As a measure of intelligence the mean Otis score of each pair was taken. Correlation coefficients between the superior effectiveness of the R-S sequence and intelligence are given in Table V.

TABLE V

COEFFICIENTS OF CORRELATION BETWEEN INTELLIGENCE AND SUPERIORITY OF THE RECITE-STUDY SEQUENCE

	r	P.E.$_r$
Mathematics classes	.115	.088
Science and English classes	.091	.126
Social Science classes	.043	.080
All classes	.063	.57

Examination of the ten correlation charts which are not reproduced here because of space limitations disclosed no significant relationship within any given range of intelligence.

CONCLUSIONS

1. It appears that neither the R-S sequence or the S-R sequence is superior for all kinds of classes. (See Table I)

2. The S-R sequence is probably superior for classes in history and social studies and literature and the superiority becomes more marked as we progress from the seventh to the eleventh grade. (See Table I)

3. The R-S sequence is probably more effective for classes in mathematics and general science, at least in the junior high schools. (See Table I)

4. Teachers' judgments as to the superior effectiveness of two methods of teaching should in all likelihood not be taken as conclusive. (See Table III) A careful study of the reliability of teachers' judgments in such instances appears very desirable.

5. Neither sequence seems to be connected in general with the production of larger variations of progress within the class than the other. (See Table IV)

6. Neither sequence seems peculiarly favorable for any given range of intelligence. (See Table V)

While no scientific conclusive explanation can be given for the above results, a study of the logs of the instructors in charge of the experimental sections and his own experience in observing and supervising supervised study technique and practice leads him to offer the following hypotheses. In those classes where the effectiveness of study does not depend so directly upon specific assignments better results will follow the use of the S-R sequence due to two factors (1) the increased interest and consequent higher level of attention in the study period immediately preceding a recitation, (2) where the S-R sequence is used the student has made a beginning of preparation, has located his difficulties and is in a receptive if not a "seeking" attitude toward assistance.

In mathematics classes, study should probably follow presentation of new material, rules, and procedures as soon as possible, while the explanations and directions are fresh in the minds of the learners. In addition the R-S sequence provides a short tryout for the assignment which often results in additional desirable explanations to the class as a whole or to individuals needing it.

Not to be ignored is the quite likely explanation that in subjects involving memory largely, recall and repetition within the next hour, as in the case of the S-R sequence, will serve to fix before the setting in of the rapid decline in the forgetting curve gets under way. In other words, study followed immediately by recall and recitation as in the S-R sequence is probably a better arrangement for acquiring history, civics, and literature facts, concepts and information, than study followed by recall and recitation 23½ hours later after the inroads of forgetting known by many previous experiments to be large have operated to wipe out much of the results of study.

EDITORIALS

SOME MATTERS OF EDITORIAL POLICY

Is the *Journal of Educational Research* too technical? Is it technical enough? Should the journal be for research specialists or should it be for teachers, principals, supervisors and superintendents interested in scientific education? Should the *Journal of Educational Research* aim definitely to modify school practice?

Dr. B. R. Buckingham, in the first issue of the *Journal of Educational Research* (January, 1920) made the following announcement of policy:

"*The Journal of Educational Research* will be devoted to the uses which have been made and are capable of being made of the results and methods of investigation. It will emphasize applications rather than abstractions, and practice rather than theory. It will not avoid the theoretical altogether, and it will present some technical articles. But such articles will be most welcome when they have obvious practical implications. Research for the sake of research we shall leave for others. What uses the teacher and supervisor have made or may make of the findings of the experimentalist, what methods are transferable from the laboratory to the classroom and office, what workable means of meeting common needs have been developed under actual school conditions—these will be our chief concern. We shall seek contributions not only from the leaders in research but also from their followers, the practical school people.

"Theory has outrun practice in our educational writings. Results are quite generally abstract, and 'tentative conclusions' are often merely speculative. Some of us wonder whether all this is necessary—whether a writer may not be scientific without being impractical or profound without being obscure—whether the time has not come when some of these results may be restated in usable form—whether we do not now possess the materials for studying some problems long enough and hard enough to reach conclusions which are fundamental and satisfactory.

"The attempt will, therefore, be to permit practice to catch up with theory. The *Journal* will endeavor to exemplify sound theory and it is hoped that its technic will be adequate. But we believe its theory and technic will be best exemplified when they are employed in the reporting of practical investigations and in the formation of practical results."

Six years later, Dr. E. J. Ashbaugh returns to the problem in an editorial, "Are We Too Highbrow?"

"For whom should educational magazines be published? What is their mission? Is their primary object simply to report or is it to report in order to modify practice? Are they publicity agencies for their contributors or are they service instruments for the benefit of the school children of the country? How much technical education have we a right to assume on the part of readers or to what extent may these magazines be expected to be the source of this technical education?

"As we examine educational magazines we find evidence of various policies. Some are so free from technical terms that it is safe to assume that anyone can read them—even university presidents, and that without a dictionary. Most of these relate personal experiences and opinions and savor little of contact with research and the scientific movement in education. They are apparently written for teachers who are neither trained nor critical. Others show evidence of an awareness of present educational tendencies. Their articles report the results of a variation of procedure which was to some extent at least checked by control methods. The language is relatively simple though the statements at times are dogmatic. Some discrimination is evidently expected of the reader though all the data are practically never presented. A third type of magazine is also available. This type fairly bristles with technical terms—quartiles, percentiles and ogives, P.E.'s and sigmas, correlations multiple and partial, and regressions. Statistics by which to validate procedure and test the results have become an important instrument in scientific educational work, and, therefore, the authors, having made use of statistical procedure, feel the necessity of reporting all of it in the reports of their experiments.

"Whenever any one develops a new formula or a new use for an old formula, technical workers will certainly be pleased to know about it. Every worker should safeguard himself by every means at his disposal in the conduct of an educational experiment and the interpretation of his findings. On the other hand, when reporting investigations which are of value to public-school people, it is certainly a mistake to burden the report with a technical vocabulary which few understand. Research findings are valuable only as they modify educational practice."

This question of editorial policy is by no means a new one. The editorial board is anxious to have the *Journal of Educational Research* serve the cause of scientific education as effectively as it can. To this end it was thought that it might be helpful to get an expression of opinion from its readers. You have already received or will receive in the near future a questionnaire asking certain questions about matters of policy. We should be glad to have you assist us by filling out this questionnaire and returning it. We shall present a digest of this material in a later issue of the *Journal*. The questions asked are:

QUESTIONS	(Check in these Columns)	
1. Should the JOURNAL OF EDUCATIONAL RESEARCH aim definitely to modify school practice? (To do so, would probably mean that the JOURNAL should attempt to serve teachers, principals, supervisors, and superintendents, who are interested in the application of scientific methods to school problems.)	Yes	No
2. Do you think that a journal of research might be more scientific and less technical? (It has been suggested that the readers want not less science, but more care exercised in making the descriptive and interpretative portions understandable to teachers, principals, supervisors, superintendents, etc., those not trained in technical research.)	Yes	No
3. As a contributor to the JOURNAL OF EDUCATIONAL RESEARCH, do you prefer to write: (1) for the research specialist alone, or (2) for teachers, principals, supervisors, and superintendents interested in scientific education, or (3) comprehensively enough to interest both groups?	(Indicate which) (1)..... (2).....	(3).....
4. Should reports of research be published which seem to follow a reasonably sound experimental technique but which are based upon a relatively small number of cases, or seem preliminary in character or not sufficiently conclusive to serve as a basis upon which to modify school practice? (A superintendent of schools recently remarked that he did not object to the reading of technical articles—as a matter of fact he preferred them—but he did object to spending an hour or more working through such an article to find in the end that it led nowhere, that is, led nowhere from his point of view.)	Yes	No
5. Is there any one field to which the JOURNAL might well give special attention during the coming year? (Such as (1) curriculum research, (2) experimental studies of teaching and learning, (3) measurements applied to character education, (4) guidance, (5) child study, (6) special education, (7) supervision, (8) administration, etc.)	Yes (Write Nos. here)	No

A. S. B.

REVIEWS

Address all communications relative to book reviews to Editor E. J. Ashbaugh, Ohio State University, Columbus, Ohio.

SYMONDS, PERCIVAL M. *Ability Standards for Standardized Achievement Tests in the High School*, New York, Teachers College, Columbia University, 1927. x + 91 pp.

Until the publication of the present monograph, both the value and the significance of the use of standardized tests in the secondary school have been seriously limited. The variability of high-school classes in mental ability has rendered the comparison of achievement test class scores with one another or with standards of little significance. Individual differences in mental ability within high-school classes has likewise handicapped any attempt to interpret the scores made by individual students.

With a grant from Teachers College and the coöperation of the principals and teachers in nearly forty secondary schools, Dr. Symonds undertook the task of determining mental ability norms for a group of standardized tests and scales for secondary schools. These mental ability norms were determined by having both the Terman Group Test of Mental Ability, and the achievement tests given to several hundred high-school students a constant unit which had been worked out previously by Thorndike. The method of equi-percentiles finally decided upon for determining the standards gives neither the average scores for pupils of varying degrees of mental ability nor the maximum expectation for gifted students but rather a series of scores that do not impose too rigid standards upon the less gifted nor too low yet attainable standards for the gifted students.

In the monograph are included tables giving mental ability standards for December 1st, March 1st, and June 1st for the American Council French Test, Haggerty Reading Test Sigma III, Hotz Equations and Problems Tests, Powers Chemistry Test, Ruch-Cossman Biology Test, Ruch-Popenoe General Science Test, Schorling-Sanford Geometry Test, Van Wagenen American History Scale S3, White Latin Tests.

For interpreting the test scores of individual pupils two methods are proposed. One simply shows the difference between the mental ability norm for the student and the student's score in terms of the test score. The second method expresses this difference in terms of the mental ability unit, thus making differences from one test to another comparable.

While the plan of Dr. Symonds assumes a constancy of units of measurement which is not yet attainable with the usual standardized test, it provides a reasonable and very valuable working basis for making a more significant use and a more adequate interpretation of achievement test scores in high-school subjects.

M. J. VAN WAGENEN
University of Minnesota

LOWMAN, C. L., COLESOCK, CLAIRE, and COOPER, HAZEL. *Corrective Physical Education for Groups.* New York, A. S. Barnes and Company, 1928. xiv + 521 pp.

No current trend in physical education is of greater significance than that of meeting the health needs of individual children but no development of the public school program could catch physical educators more unprepared to meet their opportunity Therefore, the book by Dr. Lowman and his associates should find an eager reading public, for it provides not only descriptions of corrective exercises for kyphosis, lordosis, scoliosis, foot weaknesses, etc., but also definite activity schedules adapted to groups of different degrees of maladjustment, and detailed accounts of general and specific methods of teaching.

As this volume was prepared by an orthopedic surgeon, with the aid of experienced public school physical educators specializing in corrective gymnastics, it contains much of practical value to the field worker who desires specific instructions and detailed advice in organizing and conducting an elementary type of corrective work. Unfortunately, however, the subject matter is so clumsily put together that all too often it is difficult to separate useful corrective activity material from useless general outlines. Much other material included in this volume, such as elementary physiology and kinesiology, and informational hygiene programs for physical education classes should either be related more closely to activities or be treated more comprehensively in another book.

Moreover, the chapters on organization betray a lack of clear planning or knowledge of the field of educational measurements. For example, the "Reilly system of classification" is advocated as bringing together those of equal abilities and physiological maturity. There are now several other ways of determining physical and athletic "abilities" with almost double the validity of the Reilly system. Moreover, this system, instead of bringing together those of the same physiological characteristics, tends to bring together (in the senior high school especially) most heterogeneous groups.

There is a cloud of chaff, a lack of organization of material, and some error in this volume; but the notes on equipment, the descriptions of exercises, and the corrective schedules, coupled with the great need for treatises on corrective physical education, probably justifies its publication.

FREDERICK RAND ROGERS
New York State Education Department

CRAWFORD, CLAUDE C. *The Technique of Study.* New York, Houghton Mifflin Company, 1928. 353 pp.

In *The Technique of Study*, Mr. Crawford has provided a splendid service for college students. The mass production methods of college instruction,—the stricter requirements for staying in college after one has passed the gate, the mathematical ways of grading systems, the mind-reading devices of tests and new-type "examinations," all such make demands upon students for a quality of workmanship and technical ability to handle subject matter undreamed of in an older, more human time. If a student gets into college by virtue of a high I. Q., a good high-school record, and generous character recommendations, he may stay there only by virtue of his ability to adjust himself to the college courses in a no worse than "C" equilibrium. If the balance gets

heavy toward the "D" end of the scale he is automatically down and out.

Learning how to study is usually the first and most urgent need of the freshman. The sophomore and even the junior may in their increasing wisdom perceive the value of better and speedier techniques.

Mr. Crawford does that which has made Whipple's little manual a best seller for twelve years. He arranges his book in the form of definite practical rules. He says in the preface, "A methods book should be made up largely of material which actually answers the question *how*. Material of the *why* type should be treated as next in importance, though it should be definitely related to the former type of material."

The basis for these rules is thoroughly adequate. Not only has Mr. Crawford the requisite psychological and educational background, but also he has subjected his material to careful experimental treatment. It has been tried out and revised, and tried out again in many typical situations. Nearly all the general college procedures are considered. "Taking Notes," "Listening to Lectures," "Using Textbooks," "Memorizing," and "Taking Tests" are a few of the chapter headings. Each chapter comprises a large number of rules carefully explained, a set of questions and exercises, several type tests on the content, and a bibliography.

Such a book lends itself admirably for use by a teacher as a textbook in methods of study. For students working alone without a supervisor or class the book is helpful and suggestive.

If, then, it seems necessary to mention certain inadequacies, it must not be inferred that what is missing diminishes the value of what is given. However, one can but wish that with all the opportunities for collecting material and trying it out which were at Mr. Crawford's disposal, he might have given some help in special study methods. For example, the material in the chapter on "Memorizing" makes practically no advance over James' chapter written thirty-eight years ago. Verbatim memorizing of dates may have gone out of fashion, but students who desire an A grade want to know what to remember and how to recall it at the time of need. Have history teachers, or mathematics teachers, or language teachers contributed nothing to the peculiar techniques of memorizing in their special fields? Special methods were omitted no doubt with deliberate intention, but the *how* and the *why*, as well as the monotony of presentation would have been much illuminated by the inclusion of concrete illustrations from the rich sources of the classroom.

The social aspects of study are not given sufficient space to help the student solve the extremely difficult problems presented to him daily. In most of the chapters some one of the paragraphs refers to group study, but the student is not shown how to win control over group situations either for supplementing his private enterprise or for directing or coöperating with others. Learning to "think together" should become one great objective of mass education in a democracy, an objective consciously guiding both teachers and students. The discussion method of study in college provides the habit tendencies which must work over later into our civic life and machineries of government.

Finally, Mr. Crawford scarcely tapped the reservoir of practical devices for "concentration" (a student term), for memorizing, for test-taking, indeed, for all the everyday study situations pre-

sented in every student body. "How do you do it?" is a fair question to ask of every student who maintains a high grade level. The answers reveal a wealth of practical experience applied to the specific school situation. To be sure, much of this material has not attained the dignity of "scientific" psychology, it is merely practical, it works.

The student is not given a glimpse into the pros and cons of muscle tensions, of suggestion, of fatigue and recovery, of emotion, of indirect mental vision, of atmosphere, although within these experiences one may find "those reserves of power and capacity and those latent possibilities of education and constructive thinking that are present in all of us" as Dr. Herrick so vividly points out. 'How to Study' must get deeper and farther than the orthodox, the traditional psychology of James and Whipple and Edwards and Thomas if it is to be of vital service to the student in his tremendous problem of succeeding in college.

JESSIE ALLEN CHARTERS
Ohio State University

CAVINS, L. V. *Standardization of American Poetry for School Purposes.* Chicago, The University of Chicago Press, 1928. x + 134 pp.

Dr. Cavins presents an interesting and worthwhile addition to the studies already made in classifying literature for school use. His declared purpose is to arrange poems for classroom use on the basis of their appeal to certain interests of pupils according to content. The study furthermore presents scientific evidence concerning the thought and language difficulties that children have actually found.

After furnishing a digest of the previous studies that have been made in the placement of poems in the school curriculum, the findings of a new investigation of the placement of sixty-four standard selections in the courses of study now in use in every state of the Union is presented.

The book contains a carefully analyzed symposium (by ten English scholars and textbook writers of national importance) on the "factors to be considered in placing a poem in our curriculum." It explains the construction of tests as based on the advice and experience of these experts.

The author has experimented widely in determining the validity of the tests in arranging poetry for school uses. He has furnished evidence to show that the principles upon which the tests are based are psychologically sound and in accord with accepted aims in regard to what children should derive from their study of literature.

R. E. GREEN
Ohio State University

Renaissance Student Life: The Paedologia of Petrus Mosellanus. Translated from the Latin by ROBERT FRANCIS SEYBOLT. University of Illinois Press.

Professor Seybolt has done a considerable service in translating this book and making it accessible to students of the history of education. It is a textbook which played a significant part in ushering in the Renaissance. The Paedologia of Petrus Mosellanus was first published in 1518. It answered a keenly felt human need for a lesson book from which one might learn to converse in Latin, and in the style of Cicero. Mosellanus, its author, was a distinguished humanist who belonged to the company of Reuchlin, Melancthon and Erasmus. He served with distinction as Professor of Greek at Leipzig and was rector of that university.

The conversations of this book are valuable in two respects. They tell us how Latin was studied in that educationally energetic morning of the modern world, and since they are conversations between teachers and scholars and between scholars and scholars about every day interests and school affairs, they are wonderfully informing as to the educational practices of that day. Next to the story itself nothing is so much needed for understanding the history of education as an abundance of detailed original material to make that story concrete and wholly understandable. That is the great value of this little book. Professor Seybolt has spent upon it no little skilled labor and is to be congratulated, first, upon his willingness to put himself in this fashion at the service of his fellow students of education; and second, upon the worth of the illuminating little volume which he has made. Here is one of the stairways up which our ancestors climbed in their effort to master the Latin of Cicero.

ERNEST C. MOORE
University of California at Los Angeles

RESEARCH ABSTRACTS and BIBLIOGRAPHIES

Address all communications relative to research abstracts and bibliographies to Editor A. S. Barr, University of Wisconsin, Madison, Wisconsin.

First bibliographical booklet. "Educational Literature" 1924-26. (Erstes bibliographisches Heft) (Die erziehungswissenschaftliche Forschung 1924-26).

Second bibliographical booklet. "Educational Literature" 1925-26. (Zweites bibliographisches Heft) (Die erziehungwissenschaftliche Forschung 1925-26.)

Third bibliographical booklet. "Educational Literature" "Total Bibliography of Pedagogie." (Drittes bibliographisches Heft) (Die erziehungswissenschaftliche Forschung. Paedagogische Gesamtbibliographie.)

Academy of Sciences of General Utility at Erfurt. Section for Education and Adolescence. (Akademie gemeinnuetziger Wissenschaften zu Erfurt.) (Abteilung fuer Erziehungswissenschaft und Jugendkunde.) Edited by DR. ARTHUR HOFFMAN, Publisher Kurt Stenger, Erfurt. (Verlag Kurt Stenger, Erfurt.)

The attention of American Research workers, should be called to the above publications. They offer a comprehensive bibliography of articles, treatises, dissertations, and books written in the German language dealing with education. Experts of the Academy at Erfurt and of the "National Committee on Education and Instruction" (Deutscher Ausschuss fuer Erziehung und Unterricht) at Berlin under leadership of Dr. Hoffman are responsible for these publications. By means of carefully worked out schemes of classification, keys, and other bibliographical devices, educational literature is made available. The content of the third booklet is as follows: (and it is similar to the other two): (1) German scientific educational books, 1927. (2) Bibliography of educational periodical literature. (3) German university publications, bibliography of doctor dissertations, etc. (4) Educational books in the English language, 1926. (5) Research and the teaching of education at universities and other institutions of higher learning in the German language territory, winter semester 1927. (6) The curricula at the pedagogical academies of Prussia, winter semester, 1927.

SMITH, HENRY LESTER and WRIGHT, WENDELL WILLIAM. "Second Revision of the Bibliography of Educational Measurements," *Bulletin of the School of Education*, Bloomington, Indiana, Indiana University, 1928.

This bulletin is the second revision of the Bibliography of Educational Measurements compiled by the Bureau of Co-operative Research, Indiana University, published in 1923, revised in 1925 and in 1928. The bulletin contains a complete, up-to-date, and well-arranged list of tests.

WAGNER, DR. JULIUS. "The Scientific Trend of Present Day Pedagogy." (Die Wissenschaftsmethodischen Rich-

tungen der gegen waertigen paedagogik.) *Academy of Sciences of General Utility at Erfurt, Section for Education and Adolescence.* (Academie gemeinnuetziger Wissenschaften zu Erfurt, abteilung fuer Erziehungswissenschaft und Jugenkunde) 1928.

According to Dr. Wagner, the different educational research methods are derived from three mental attitudes: the empirical, the speculative, and the critical. The following topics are discussed under empirical pedagogy: 1. The direction of empirical pedagogy. 2. The empirical method in pedagogy. 3. The experimental procedure. (A) The didactical and the psychological empiricism, (a) Foundation by Lay and Meumann, (b) Pedagogical psychology and its problems. (B) Biological empiricism, (C) Sociological empiricism.

Under critical pedagogy we find the following titles: 1. The nature of criticism. 2. Methodological problem. 3. Phenomenological and the descriptive method. 4. Normative pedagogy. (a) Pedagogical teleology. (b) The value-problem in pedagogy. (c) Task and significance of the pedagogical theory of values. (d) The pedagogical norm-sciences. Neither empirical, speculative nor critical methods of attacking educational problems can lead to a complete final system of education according to Dr. Wagner. They supplement each other and must in turn be supplemented by the philosophical method.

WEIDEL, KARL, PROF. "The New Teacher Training in Prussia" (Die neue Lehrerbildung in Preussen) No. 13 of the publication of the *Academy of sciences of general utility at Erfurt* (Akademie gemeinnuetziger Wissenschaften zu Erfurt) *Section for Education and Adolescence* (Abteilung fuer Erziehungswissenschaft und Jugendkunde). 1928, Publisher: Verlag Kurt Stenger, Erfurt.

The first attempt to reorganize teacher training was made by Prussia in establishing new pedagogical academies at Bonn, Kiel, Frankfurt a. M., and Elbing. Teachers have first to graduate from a gymnasia, then visit for two years a pedagogical academy and pass the first teacher examination. After teaching for four years they have to pass the second teacher examination. Hereafter they can be appointed as state officer.

The course of study, aims and objectives, methods and procedures, etc., are fully discussed by Professor Weidel, as academy director at Elbing. An appendix contains the actual course of study for the last year, the new regulation for the first and second teacher examination in Prussia, the equipment of the academy, etc.

EXPERIMENTAL STUDIES REPORTED IN CURRENT ISSUES OF OTHER MAGAZINES

PSYCHOLOGY OF LEARNING

BARLOW, M. C. "The Role of Articulation in Memorizing," *Journal of Experimental Psychology,* XI (August, 1928), 306-312.

A study of the advantage, if any, of articulation in memorization. The subject-matter consisted of two 20-word lists of three-letter syllables that conveyed no obvious meaning. Three hundred and twenty-seven subjects ranging in school from the second to the ninth, in addition to 51 students of college level were used. The results are briefly summarized by the author as follows: (1) The average result from every one

of our groups shows that the syllables were more economically learned with full than restricted articulation. (2) The girls make better scores than the boys in both kinds of learning. (3) The pronounced individual differences in the relative number of syllables learned by the two methods point to a need of experimental determination for each pupil of the better method for his use in memorization. (4) The errors made by the first-described method are slightly in excess of those made by the latter. They are about equally numerous in the lower and in the upper grades, and more numerous in the intermediate grades.

BRUECKNER, LEO J. "Analysis of Difficulties in Decimals," *The Elementary School Journal*, XXIX (Sept. 1928), 32-41.

This article contains the results of a study of the errors made by pupils in Grades VI, VII, and VIII in analytical tests on the four processes in decimals. The basis of the analysis was a study of the written work of some three hundred pupils in these grades. The writer was primarily concerned with the types of examples found most difficult and the causes of error. Tests were given (these are reproduced in the article) and the incorrect examples examined to determine causes of error. The total number of different kinds of errors isolated was 114; the kinds of errors were more numerous in multiplication and division than in addition and subtraction. They were distributed as follows: general difficulties in reading, writing, and conversion, 33; difficulties in addition, 15; difficulties in subtraction, 14; difficulties in multiplication, 26; and difficulties in division, 26. The tables reproduced in the article contain the classification of 8,785 errors. The author draws the following conclusions:

1. Many pupils do not have adequate concepts of the numerical values of decimals. Only two other types of errors occurred as frequently, namely, misplacing the decimal point in division and misplacing the decimal point in multiplication.

2. Many errors were due to the misspelling of the decimal written in word form; for example, "hundreds" for "hundredths."

3. Failure to place the decimal point correctly was the greatest cause of errors in addition, $.3 + .5 + .8 = .16$ being the most common type.

4. The number of errors in addition due to inaccuracy was about half as great as the number of errors due to the misplacing of the decimal point.

5. The greatest difficulties in subtraction were in borrowing and in the placement of the decimal number in the subtrahend. There were few errors due to inaccuracy.

6. The major difficulty in the multiplication of decimals was the misplacing of the decimal point or its complete omission.

7. There were many errors due to inaccuracy in multiplication.

8. The major causes of errors in division were the misplacing of the decimal point, faulty placement of zeros, omission of the decimal point, and inaccuracy.

FENNELL, ETHEL L. "Recurring Words and their Relations to Difficulties in Comprehension," *The Elementary School Journal*, XXIX (Sept. 1928), 42-53.

The study of recurring words of different meanings, found in readers. The article points out that a probable reason for difficulty in the comprehension of reading lies in the fact that textbook writers in their effort to familiarize

children with certain words have provided for numerous recurrences of these words but have disregarded the fact that frequently the words recur with entirely different meanings.

PARSCHAU, FRANZ. "Testing the Permanency of School Knowledge." (Pruefung des Dauererfolges der Schularbeit.) Zeitschrift fuer paedagogische Psychologie, experimentelle Paedagogik und jugendkundliche Forschung. (July-August, 1928), 351-356.

Investigations of the knowledge and skill of young school graduates in industry, a number of investigations are referred to, two of which are summarized below:

One was made at the University of Jena by H. Meister on "The Retention of School Knowledge in Its Relation to the Personality." (Die Retention des Schulwissens in ihrer Beziehung zur Persoenlichkeit), and another at the University of Koenigsberg by Mueller-Jung on "School Knowledge and Skill by Graduates." (Schulmaessiges Wissen und Koennen bei Schulentlassenen.)

Mr. Meister examined 20 male persons, 20-25 years of age. The time since leaving school was from 6 to 11 years. He put 162 oral questions on all elementary subjects to them, and evaluated the answers from 0 to 3 points using the best possible answer as criteria He found that they retained in Arithmetic 57 percent, in German 61 percent, in History 57 percent, in Geography 62 percent, in Natural Science 52 percent, and in Religion 32 percent of the subject matter taught. Thus there remained about 50 percent of the facts taught in elementary grades.

Mueller-Jung examined 100 girls attending vocational schools who had left elementary school 1 to 3 years before. He secured the following results: Arithmetic 54 percent, History 35 percent, Geography 47.5 percent and Natural Sciences 56 percent. The average on the later mentioned subjects was 48.1 percent. Both investigations conclude that graduates of the elementary school retained about 50 percent of the facts that had been taught them. These findings open the elementary school to criticism in regard to the efficiency with which facts are taught.

TINKER, MILES A. "Eye Movement Duration, Pause Duration, and Reading Time," *Psychological Review*, XXXV (September, 1928), 385-397.

The purpose of this study was: (1) To furnish new data on the extent and duration of eye movements in reading; (2) To show the variation in the ratio between eye movement duration and total reading time in various kinds of material. A photographic technique was employed to record eye movements. Four kinds of material were used: (a) Thirteen lines of easy scientific prose; (b) four paragraphs of elementary Chemistry; (c) Two paragraphs of Algebra; and (d) Four lines of Algebraic formulae. The authors concluded:

1. That his study yielded reading movements to the right which varied from 1° (formulae) to 2° (prose); movements to the left being from 13° (prose) to 16° (formulae).

2. For all materials, less than 10 percent of the reading time was taken by eye movements. The total average time taken for eye movements was 5.9 percent of the reading time. This allows about 94 percent for perception, i. e., pauses.

3. The ratio between eye movement duration and reading time bear a constant relationship to the reading at-

titude. It was found that the more careful and analytical the reading the smaller this ratio, when the mental reading processes involved in the reading were more complex the reading pause consumed a relatively greater percent of the total time.

TINKER, MILES A. AND PATERSON, DONALD G. "Influence of Type Form on Speed of Reading," *The Journal of Applied Psychology*, XII, (August, 1928), 359-368.

Forms A and B of the Chapman-Cook Speed of Reading Test were set in 10 point old style in each of the three forms, namely, all capitals, lower case, and italics. Eight groups of 80 each were tested, or 640 in all. The results are summarized by the authors as follows:

1. A speed of reading record was obtained for texts set up in roman lower case, italic, and all capital letters. Comparisons were made between all capital versus lower case text. There were 640 subjects used in the experiment.

2. The text in lower case letters was read 13.4 percent faster than that in all capitals. The difference in favor of the lower case text is highly significant. There were 4.74 words read per second of the all capitals and 5.38 words per second of the lower case text.

3. The text in lower case letters was read 2.8 percent faster than the italics. The direction of the difference was consistent for all sub-groups. The subjects read 5.08 words per second of the italics and 5.22 words per second of the lower case text.

4. The following conclusions seem justified by these data: (a) Text printed in all capitals is less legible than material in lower case letters and slows down speed of reading to a very marked degree. (b) Text printed in italics is slightly less legible than material in lower case letters and decreases speed of reading somewhat.

MISCELLANEOUS

BECKMANN, DR. H. "The Composition as a Means of Selecting Pupils". (Der Aufsatz im Dienste der Schuelerauslese). Zeitschrift fuer paedagogische Psychologie, experimentelle Paedagogik und jugendkundliche Forschung. XXIX (April 1928), 199-205.

From a previous study of the grammatical development of the written expression of pupils of grade-school age the author developed a scale which he uses in this investigation. This scale which is a scale showing the frequency with which different parts of speech are used in different grades is given below: (reported on the above Journal No. 6, 1927, p. 274)

Grades	Abstract Noun	Adjective	Adverb	Preposition	Relative Pronoun
1	0.0%	1.1%	17.7%	4.9%	0.3%
2	1.1%	2.5%	15.7%	5.0%	0.6%
3	1.2%	2.7%	12.2%	6.8%	0.9%
4	2.1%	3.6%	8.6%	7.0%	1.5%
5	2.7%	3.7%	9.4%	7.6%	1.6%
6	4.7%	4.2%	7.3%	8.5%	2.2%
7	5.5%	4.8%	5.9%	10.0%	3.3%
8	5.5%	5.8%	4.4%	10.6%	3.9%

This table indicates that 0.0 percent of the written language of the pupil of the first grade consists of abstract nouns; 1.1 percent consists of adjectives, 17.7 percent consists of adverbs, etc. In the investigation reported herein 233 fourth-grade pupils wrote a composition which Mr. Beckmann evaluated by first determining the frequency of occurrence of the parts of speech and then by comparing the findings with norm for the fourth grade as given in the above scale.

The percents ranged from minus 12.4 to plus 10.7 percent from the average for the fourth grade which was plus 4.56 percent. The rank order of the occurrence of these parts of speech was then compared with marks assigned to compositions and also with marks in German and other subjects. Correlations between .38 and .46 were found. These fourth-grade pupils later went to the secondary school. After one year of work their standing in the higher school was compared with their standing in the elementary school and the author found that pupils who scored from plus 4.1 percent to 10.7 percent (or who were above the average, 4.56 percent) were 100 percent successful in the higher school, that of those who scored from plus 4.1 percent to plus .8 percent only 74.6 percent succeeded, from plus .8 percent to minus 2.5 percent only 64.8 percent succeeded, from minus 2.5 percent to 5.8 percent only 55.6 percent, from minus 5.8 percent to 9.1 percent only 62.4 percent and from minus 9.1 percent to 12.4 percent only 42.9 percent succeeded. The author sees in these data a close correlation between proficiency in language and success in school work.

The Journal of Educational Psychology, XIX, (September, 1928) carries an excellent discussion of nature and nurture and their influence upon intelligence and school achievement. Seven papers which formed the nucleus of the February, 1928 program of The National Society for the Study of Education, are printed in this issue of the *Journal*. These papers are:

(1) Terman, Lewis M. "The Influence of Nature and Nurture upon Intelligence Scores: An Evaluation of the Evidence in Part I of the 1928 Yearbook of the National Society for the Study of Education." (2) Terman, Lewis M. "Replying to Criticisms of Part I, and Introducing Part II of the 1928 Yearbook of the National Society for the Study of Education." (3) Freeman, Frank N. "An Evaluation of the Evidence in Part I of the Yearbook and Its Bearing on the Interpretation of Intelligence Tests." (4) Gates, Arthur I, "Observed Facts and Theoretical Concepts." (5) Whipple, Guy M. "Editorial Impression of the Contribution to Knowledge of The Twenty-Seventh Yearbook." (6) Judd, Charles, H. "Intelligence as Method of Adaptation." (7) Baldwin, Bird T. "Heredity and Environment—or Capacity and Training."

NEW TESTS

Classroom Procedure Test. Waples and Reavis. University of Chicago Press, 1928.

Classroom Procedure Test in Mathematics. Breslich and Stone. University of Chicago Press, 1928.

Classroom Procedure Test in Natural Science. Cunningham and Waples. University of Chicago Press, 1928.

Classroom Procedure Test in History and Other Social Studies. Wilson and Waples. University of Chicago Press, 1928.

Classroom Procedure Test in English Composition. Anderson and Waples. University of Chicago Press, 1928.

Classroom Procedure Test in English Literature. Poley and Waples. University of Chicago Press, 1928.

The Standard High School Spelling Scale, (Revised and Enlarged Edition). Ernest P. Simmons and Harold H. Bixler. Smith, Hammond & Co., Atlanta, Georgia, 1928.

Ullman-Clark Test of Classical References and Allusions. Bureau of Educational Research, University of Iowa, 1928.

NEWS and COMMUNICATIONS

Address all research news and communications to Editor Earl Hudelson, College of Education, University of Minnesota, Minneapolis, Minnesota.

The National Association of Commercial Teacher-Training Institutions announces the initiation of a program of coöperative research leading to the construction of a functional and effective commercial teacher-training curriculum. Professor F. J. Weersing of the University of Southern California, chairman of the committee, is now soliciting suggestions for such a program of research.

The many administrators who by their annual reports show that they are vitally concerned with the effect of attendance upon pupil progress will be interested to learn that Dr. Carl W. Ziegler has evolved a concrete measure of school efficiency in terms of attendance and progress. His "age-delay" scheme affords parents a quantitative measure of the importance of regular pupil attendance. The report is published by Teachers College, Columbia University.

An intensive study of the Kindergarten-6-4-4 plan of organization is under way in Pasadena, California, under the direction of Supt. John A. Sexson. The first Bulletin report, entitled *Some Fundamental Concepts Underlying the 6-4-4 Plan of School Organization*, is already published. Other Bulletins announced for the current year are: *Administrative Organization of the Four-Year College, Student Government of a Four-Year Junior College, The Economics of the 6-4-4 Plan, The Secondary Curriculum under the 6-4-4 Plan*, and *The Guidance Program of the Four-Year Junior College*. Copies of these reports can be had by addressing the Principal's Office, Pasadena Junior College.

An example has just come to hand of a measuring instrument devised in the light of the results of previous investigations into what it is important for elementary-school pupils to know about the geography of the world. It is the *Witham Geography Test*, Forms I and II, published by the Research Service Company, Los Angeles. They are distinctly information tests; and, while they for the most part deal with such unrelated facts as location, size, products, etc., they give some attention to the social-economic significance of isolated facts. The selection of content in the tests is based mainly upon the findings of Rugg and Hockett.

In the belief that teacher training will probably never be consummated in educational institutions and with ample precedents in other professional fields, the University of Wisconsin has initiated a coöperative project of continued training of teachers in service. The general aspects of the work will be handled through the Supervisory Service Bureau under the direction of Professor A. S. Barr and Assistant Director T. L. Torgerson. Allocation of special services is indicated in the following announcement:

The bureau which is established to assist inexperienced teachers, graduated from the University of Wisconsin and teaching in the State of Wisconsin, in making the transition from university work to classroom teaching, proposes to render the following services:

1. A correspondence service for teachers, principals, and superintendents asking for assistance.
2. Visitation on call to teachers, principals, and superintendents asking for assistance.
3. Bulletin service dealing with specific teaching problems, the teaching of various school subjects, and the supervision of instruction.
4. Research service upon the training of teachers in service, including a continuous survey of the teaching difficulties, errors, and shortcomings of beginning teachers, graduates from the University of Wisconsin, studies of means of improvement and the improvability of beginning teachers, tests of teaching efficiency, etc.

Plans have been completed for a new twelve-story School of Education building at the Washington Square center of New York University. The significance of this to educational research is shown by the fact that an entire floor of the new building will be devoted to research in physical education under the direction of Professors Jay B. Nash, S. U. Lawton, and C. W. Hetherington. Equally spacious accommodations will be afforded the departments of music education, vocational education, art education commercial education, educational sociology, and educational psychology. The new building is a necessary result of the growth of enrollment in the school of Education from less than 150 to more than 4,000 during Dr. John W. Withers' seven years as Dean.

Does the number of pupils electing a subject tend to affect the percent of failures in that subject? The question is prompted by a recent report from Supt. M. G. Clark and a Research Committee of the Administration Club of Sioux City, Iowa. The purpose of the report is to interest all teachers in the causes and elimination of pupil failures. The committee recommends that the study be continued during 1928-29. It appears from the data for the past year that the highest percents of failure were in populous required courses and the lowest percents were in small elective departments.

Reprints of the report of the Special Committee on the Place of Science in Education may now be had from the chairman, Dr. Otis W. Caldwell, Institute of School Experimentation, Columbia University, or from the vice-chairman, Dr. Elliot R. Downing, University of Chicago.

More and better educative material and instructional devices based upon the results of curriculum research are constantly becoming available. One bigger and better example that has just come to hand is the revised and enlarged edition of *The Standard High School Spelling Scale*, by Superintendent Ernest P. Simmons of Elliott, Iowa, and Dr. Harold H. Bixler, Director of Testing and Guidance, Atlanta. It is an admirable example of the collation of results of previous investigations and the application of coöperative research to classroom problems.

An attempt to effect testing economy has been made by C. C. Peters, Director of Educational Research at Pennsylvania State College, who has devised a cultural background test that can be used over and over. All marks are made on a loose strip of blank paper. In form

the test follows the suggestion of Dr. Greene in the February, 1926, issue of the *Journal of Educational Research*. The items are grouped according to their social importance. The test is diagnostic and it is being standardized. It is announced as the Peters-Lozo Test of General Information.

During the current wave of interest in character traits among pupils, it is to be expected that existing devices for measuring personal attributes will be subjected to renewed inspection. Dr. Richard S. Uhrbrook has recently scrutinized the Downey Will-Temperament Tests. His analyzed results appear as Contributions to Education, No. 296, and are published by Teachers College, Columbia University. The report is a discriminating evaluation of the merits and inadequacies of the original tests.

Many schools far and wide are reporting that the Public School Achievement Tests have been received and that the Fifth Annual Nation-Wide Testing Survey is under way. The tests are designed for grades 3 to 8 and include reading, arithmetic computation and reasoning, language usage, spelling, grammar, history, geography, nature study, and health. Administrators and supervisors report that the annual survey is a spur to both pupils and teachers and that the diagnostic tests are lifting supervision above the level of guesswork.

The first occupant of the newly created William L. Honnold Lectureship at Knox College will be Miss Ida M. Tarbell who will lecture for six weeks on *The Growth of Big Business in America*. In addition, Miss Tarbell will conduct a series of informal conferences for women on the vocational problems of college women.

The report of the recent school building survey of Paducah, Kentucky, is now availible from the Bureau of Publications, Teachers College. The survey, made by Professors George D. Strayer and N. L. Englehardt proposes a feasible building program for the immediate future and recommends a way of meeting the financial obligations of such a program. L. J. Hanifan is superintendent of schools at Paducah.

Professor E. A. Kirkpatrick retires this year after thirty years of service in the State Normal School at Fitchburg, Massachusetts. He is now revising his well-known work on *Child Study* and will continue to pursue his interest in the ethical aspects of sociology.

Dr. Marvin F. Beeson, formerly Director of Education in State Teachers College, San Diego, California, has been appointed Associate Professor of Educational Psychology and Education at the College of William and Mary in Williamsburg, Virginia.

Dr. Harl R. Douglass of the University of Oregon is spending a year's leave of absence as visiting lecturer in Secondary Education at the University of Pennsylvania.

THE EFFECTS OF DRILL IN ALGEBRA

During the year 1925-26 the writer made a study of the effects of drill on a special class in algebra. The practice material was the accuracy drill tests of Schorling, published by the World Book Company. Two classes were used, one as the experimental class and the other as a control class. They were not taught by the same teacher, but were given the same material in algebra, with the exception that the experimental group was given eight-minute drill periods from

three to five times a week for seven months.

The achievement of the classes in algebra was measured by the Holtz Tests. The short form of the equation and addition tests was given in October and the long form of the equation, addition, multiplication and problem tests was given in May.

A third, normal group, taught by another teacher, was also given the Holtz Tests, but had no special drill.

In chronological age, the experimental class averaged 186.8 months, with S.D. 26.6 months. The control group had a mean age of 174.6 months, with S.D. 12.5 months. The normal group had a mean age of 175.2, S.D. 9.31 months. The average age for all groups combined was 184 months, S.D. 24.99 months.

The age of the experimental group was more than a year above that of the other classes, and it had an S.D. more than twice as great as the control class and nearly three times as large as the normal algebra group.

The I.Q.'s of the experimental and control groups were paired to assure equivalent mental groups. The experimental group had a mean I.Q. of 98.8, S.D. 16.4. The control group averaged 98.6, S.D. 14.6. The mean I.Q. of the normal group was 110.4, S.D. 10. The average I.Q. of all groups combined was 99, S.D. 16.26. The McCall Multimental Test of Intelligence was used in this study.

Groups 1 and 2 were ten to fifteen points lower in intelligence than Thorndike estimates a student should be to study algebra with profit. The normal group was near the estimated standard for success in algebra.

Groups 1 and 2 were somewhat retarded in grade-school achievement as measured by the Stanford Achievement Test. The experimental group had a mean E.Q. of 90.7, S.D. 13.6. The mean E.Q. of the control group was 94, S.D. 7.6. The normal group had a mean E.Q. of 106.5, S.D. 8.76. The mean E.Q. for all groups was 93, S.D. 13.8.

The various A.Q.'s or A.R.'s for the classes were determined by changing Mental Ages, Total Algebra Scores, Teachers' Grades in Other Subjects, and Educational Ages into T scores and dividing the various T scores by the Mental Age T Score. All Algebra Test Scores were totaled without weighting various test scores.

The experimental group had a mean total Algebra A.R. of 91.5, S.D. 16.2. The control group had a mean total Algebra A.R. of 95.3, S.D. 26.2. The normal group had a mean total Algebra A.R. of 112, S.D. 4.34. All groups had a mean of 102.6, S.D. 25.47.

The experimental group's mean A.R. of E.A. was 95.3, S.D. 16.2. The control group's mean was 101, S.D. 25.7. The normal group's mean was 108.7, S.D. 22.56. Total mean, 102.5, S.D. 23.74.

The experimental group's mean A.R. of Teachers' Grades in Algebra was 88, S.D. 15.1. The control group's mean was 105, S.D. 31.3. The normal group's mean was 100, S.D. 11.8. The total mean was 103.2, S.D. 25.83.

The experimental group's mean A.R. of Teachers' Average Grade for subjects other than Algebra was 91, S.D. 20.7. Control group's mean was 102, S.D. 14.4. The normal group's mean was 98.8, S.D. 11.9. Total mean, 102.3, S.D. 26.66. Mean Algebra Scores for the four groups were 27.3, 23.2, 50.4, and 32.6, with S.D.'s of 6.76, 9, 8.66, and 13.27.

RESULTS

The experimental group had a mean total score of 27.3., S.D. 9, plus or

minus 1.7; the control group had a mean total score of 23.2—a difference of 4.1. The experimental coefficient was .46. The chances are 89 in 100 that the true difference between the experimental and control group means is greater than zero. By the T Score the difference was 3.5 and E.C. was .45.

The experimental group had a mean score on the addition test of 8.2, S.D. 4.16. The control group had a mean score on the same test of 6.2, S.D. 2.6. The difference was 2, and the chances 94 in 100 that the true difference is greater than zero.

The variability coefficients were 50.7 for the experimental group and 43.2 for the control group.

In the multiplication and division test the means were 9.6 and 8.6; S.D.'s of 2.9 and 3.3. The difference in favor of the experimental group was 1, and the chances 80 in 100. The coefficients of variability were 30.3 and 38.3.

In the equation and formula test the means were 6.6 and 5.7, with S.D.'s of 2.8 and 2.9; and the difference in favor of the experimental group was .9. The chances are 80 in 100 and the coefficients of variability 42.4 and 50.9.

The problem tests showed means of 2.8 for both groups.

As measured by the gain, the experimental group showed a 12 percent advantage over the control group in total score. Since the greatest gains were in addition and multiplication the advantage seems to have been the result of drill.

In effort, as measured by accomplishment ratio, the experimental and control groups show constancy in algebra, other high-school work, and grammar-school work, none of the A.R.'s varying more than 5 percent above or below the other. The normal group also showed steadiness of effort.

The experimental group had a coefficient of variability in total score of 32.9, while the control group had a coefficient of 29.1.

BYRON K. ARMSTRONG
Lincoln High School, Kansas City